Abstract

Efficient genome engineering in mammalian systems using TAL effector nucleases and CRISPR/Cas system

Duk Hyoung Kim

Department of Chemistry

The Graduate School

Seoul National University

Programmable nucleases like transcription activator-like effector nucleases (TALENs) and RNA-guided engineered nucleases (RGENs) are broadly used for engineering of genome in various cells and organisms. TALENs are recently developed fusion proteins of a modular DNA-binding TALE repeat domain and a catalytic nuclease domain of FokI restriction endonuclease. In this thesis, I designed and synthesized highly active TALENs to target the *progesterone immunomodulatory binding*

factor 1 (*Pibf1*) and the *selenoprotein W, muscle 1* (*Sepw1*) gene. Non-homologous end-joining (NHEJ) mediated indel mutations were detected in transfected mouse cells with plasmids encoding these TALENs and knock-out mice of these genes were successfully generated via embryo microinjection. RGENs are efficient artificial endonucleases using CRISPR/Cas system which is a prokaryotic immune system. In this study, I have designed RGENs to target inverted region of the blood coagulation *factor VIII* (*F8*) gene that is known as a major cause of severe hemophilia phenotype. When two RGENs were used to generate double-strand breaks (DSB) at their endogenous target loci in wild-type Hela cells, the large chromosomal segment up to 600-kbp is inverted. Furthermore, the inverted region of induced pluripotent stem cells (iPSC) from hemophilia patients is also successfully reverted via these RGENs and mRNA expression of *F8* is recovered. TALENs and RGENs have enabled targeted genome editing easily and efficiently. These technologies will be very useful systems for research of gene functions and correcting the genetic causes of many diseases.

Keywords : Genome engineering, Transcription activator-like effector nuclease (TALEN), RNA-guided engineered nuclease (RGEN), Clustered Regularly Interspaced Short Palindromic Repeats(CRISPR), Engineered nucleases

Student Number : 2008-22716

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AP	alkaline phosphatase
Cas	CRISPR-associated
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DSB	double strand break
fPCR	Fluorescence PCR
F8	coagulation factor VIII
HA	Hemophilia A
iPSC	induced pluripotent sten cell
NHEJ	Non-homologous end-joining
PAM	Protospacer Adjacent Motif
Pibf1	progesterone immunomodulatory binding factor 1
RGEN	RNA guided engineered nuclease
RNP	ribonucleoprotein
Sepw1	selenoprotein W, muscle 1
TALEN	Transcription activator-like effector nuclease
T7E1	T7 Endonuclease 1
ZFN	Zinc finger nuclease

Chapter 1. Efficient TALEN-mediated Gene Targeting in Mice

Introduction

Elucidating gene function and identifying gene products are the great interest of many scientists. Although human and mouse genome have been sequenced, the network and function of most genes are still not uncovered.

Specific target gene disruption is a powerful method of elucidating gene function *in vivo*. For example, targeted gene knockdown by RNA interference (RNAi) is a rapid, inexpensive, and high-throughput system for reverse genetic approaches. RNAi was first discovered in the worm (Fire et al. 1998) and many researchers have recently reported successful siRNA-mediated knockdown of mammalian genes (Elbashir, S. M. et al. 2001; Ohta, T. et al. 2002; Moskalenko, S. et al. 2002). However, RNAi is incomplete because of temporarily knockdown rather than gene knockout, varies between experiments and laboratories, and even has unpredictable off-target effects (Krueger et al. 2007; Jackson et al. 2003).

Introducing a double-strand break (DSB) in a eukaryotic chromosomal DNA can stimulate nonhomologous end-joining (NHEJ) repair system and produce localized indel mutations at the break (Jeggo 1998; van Gent et al. 2001) which can lead to gene disruption. Thus, targeted cleavage of specific DNA sequence *in vivo* is a very powerful research technique for studying genome function. Zinc finger nucleases

(ZFNs) are engineered nucleases generated by fusing a zinc finger DNA-binding domain and a DNA-cleavage domain of FokI restriction enzyme (Kim et al, 1996). Because zinc finger DNA-binding domains can be designed to target a specific sequence, introducing genetic modifications in a target gene can be possible. Actually, targeted genome editing is successfully performed in mammalian, plant, and other higher eukaryotic cells and organisms (Bibivoka et al. 2002, 2003; Lloyd et al. 2005; Urnov et al. 2005; Wright et al. 2006; Beumer et al. 2006; Doyon et al. 2008; Maeder et al. 2008; Meng et al. 2008; Santiago et al. 2008)

Although ZFNs are very useful tools for gene editing, there are some complicating issues about the design and application of ZFNs. First, it is difficult in some cases to design ZFNs that have the binding specificities at the desired target sites (Ramirez et al. 2008). A zinc finger protein can recognize three base pairs specifically, but assembled zinc finger proteins into arrays do not always recognize desired sequences (Lam et al. 2011). Second, some ZFNs have cytotoxicity, presumably due to making DSB at off-target sites (Cornu et al. 2008; Radecke et al. 2010; Pruett et al. 2008). Although this problem has been solved to some extent with using engineered FokI heterodimers (Szczepek et al. 2007; Miller et al. 2007; Doyon et al. 2011), off-target effect of ZFNs is a problem that is still concern.

Transcription activator-like effectors (TALEs) are natural modular proteins secreted by pathogens of plants like *Xanthomonas* to

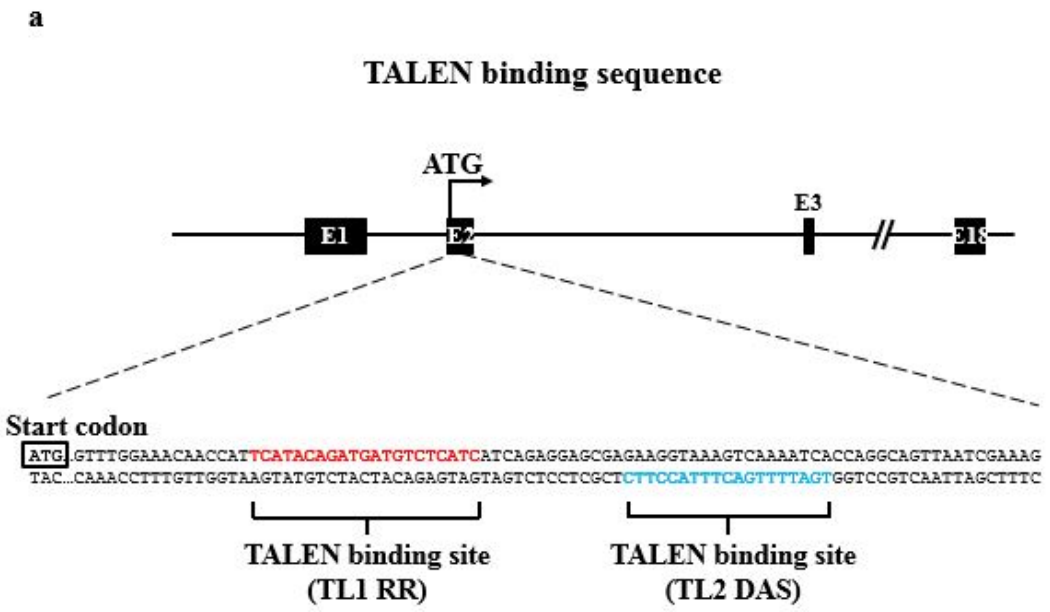
modulate gene expression in host plants and to facilitate bacterial survival. (Boch et al. 2010; Bogdanove et al. 2010) Repeat monomers of TALEs differ from each other mainly in amino acid positions 12 and 13, and a strong correlation have been revealed between pairs of amino acids at positions 12 and 13 and the corresponding DNA nucleotide in the TALE-binding site: NI to A, HD to C, NG to T, NN to G (Boch et al. 2009). Thus, like ZFN, TALEN which is a novel fusion protein of a TALE protein and a DNA-cleavage domain of FokI restriction enzyme enables the targeted alteration at a predetermined locus (Cermak et al. 2011; Miller et al. 2011; Sander et al. 2011).

Engineered nuclease-mediated mutagenesis in zygotes is a potential alternative to conventional gene targeting in mice. Although microinjection of engineered ZFNs in embryos was used to generate gene knockouts in the mouse successfully (Carbery, I.D. et al. 2010), gene knockout mice have yet to be created using TALENs. In this thesis, I designed and synthesized highly active TALENs to target the *progesterone immunomodulatory binding factor 1* (*Pibf1*) gene and the *selenoprotein W, muscle 1* (*Sepw1*) gene in mice. In NIH3T3 mouse cell line, TALEN activity is confirmed by T7 Endonuclease 1 (T7E1) assay and sequencing and knock-out mice of these genes were successfully generated via embryo microinjection of each TALENs.

Materials and Methods

1. Plasmid encoding TALENs

For Pibf1-specific TALEN plasmids, oligonucleotides encoding each TALE repeat module were synthesized and subcloned into modularly assembled repeat arrays, with the ⊿153/+16 N- and C-terminal truncation domains of AvrBs3 and the FokI nuclease domain. FokI nuclease domains of TALENs used in this study were the obligatory heterodimer nucleases Sharkey RR and Sharkey DAS (Guo et al. 2010). The Sepw1-specific TALEN plasmids were obtained from ToolGen, Inc. (South Korea). The complete sequences of the TALENs targeting *Pibf1* and *Sepw1*, and their respective target sequences, are shown in Figure 1 and 2.



(continued)

Figure 1. Target sequences of Pibf1−TALEN. **(a)** The target region of Pibf1−TALEN in the mouse Pibf1 locus. Pibf1−TALEN binding sites: red, TL1−RR; blue, TL2−DAS. **(b and c)** Nucleotide and amino−acid sequences of TL1−RR **(b)** and TL2−DAS **(c)**.

b

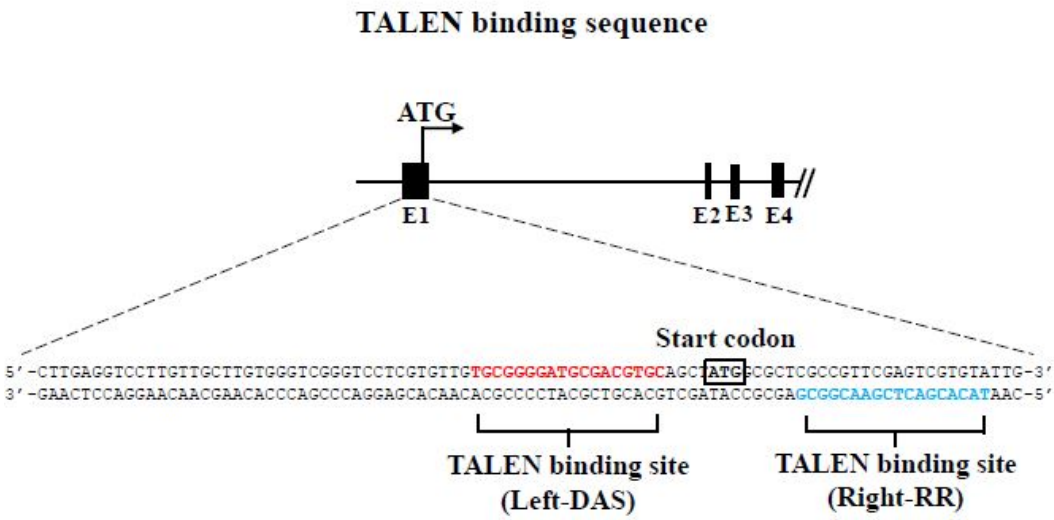

>TL1-RR
HA Tag
NLS
TALEN module
+16 C-terminal domain of AvrBs3
TAG 3123
FokI nuclease domain

(continued)

c


>TL2-DAS
HA Tag
NLS
TALEN module
+16 C-terminal domain of AvrBs3
FokI nuclease domain

a



(continued)

Figure 2. Target sequences of Sepw1-TALEN. **(a)** The target region of Sepw1-TALEN in the mouse Sepw1 locus. Sepw1-TALEN binding sites: red, Left-DAS; blue, Right-RR. **(b and c)** Nucleotide and amino-acid sequences of Left-DAS **(b)** and Right-RR **(c)**.

b


>Left-DAS
HA Tag
NLS
TALEN module
+16 C-terminal domain of AvrBs3
FokI nuclease domain

(continued)

c


>Right-RR
HA Tag
NLS
TALEN module
+16 C-terminal domain of AvrBs3
FokI nuclease domain

2. Non-homologous end-joining (NHEJ) reporter assay

NHEJ reporter plasmids used in this study were constructed as previously described (Kim et al. 2011). In brief, oligonucleotides containing each target site were synthesized (Bioneer, Inc., Korea), and the surrogate reporters were constructed by cloning annealed oligonucleotides into vector pRGS (ToolGen, Inc., Korea). Each surrogate reporter was transfected into HEK-293 cells with or without TALEN-expressing constructs using Lipofectamine 2000 (Invitrogen), and the reporter-gene activities were measured by flow cytometry 3 days after transfection using a FACSCanto flow cytometer (BD Bioscience) or C6 flow cytometer (Accuri).

3. Transient transfection and T7 endonuclease I (T7E1) assay

NIH3T3 cells were transfected with TALEN pairs using Turbofect (Fermentas) according to the manufacturer's instructions. The genomic DNA samples were extracted 72 hours after transfection using the Genomic DNA Extraction Kit (iNtRON Biotechnology) as described by the manufacturer. To screen F0 mice with targeted *Pibf1* or *Sepw1* alleles, genomic DNA samples were prepared from tail biopsies. The T7E1 assays were performed as previously described (Kim et al.

2009). Briefly, the genomic region encompassing the TALEN target site was PCR-amplified, melted, and re-annealed to form heteroduplex DNA, which was treated with 5 units of T7 endonuclease 1 (New England Biolabs) for 15 min at 37° C and then analyzed by agarose gel electrophoresis. To evaluate non-specific effects of TALENs, potential off target sites were predicted as previously described (Doyle et al. 2012), and T7E1 assays were performed using genomic DNA samples from F0 mutant mice.

4. Sequence analyses and genotyping

To confirm the genotype of F0 mice detected by T7E1 assays, further analyses were conducted as follows. For fluorescence PCR (fPCR), genomic DNA was subjected to PCR analysis using HiPi Plus DNA polymerase (ELPIS) and 5′ -carboxyfluorescein-labeled primers. The PCR amplicons were studied using an ABI 3730xl DNA analyzer. For sequencing, PCR products were cloned using the T-Blunt PCR Cloning Kit (SolGent Co., Ltd. Korea), and mutations were identified by direct sequencing. For routine PCR genotyping of F1 progeny from selected F0 founder mice, the following primer pair was designed to amplify an 80-bp PCR product from wild-type mice and smaller products from TALEN-induced mutant alleles: 5'- GTTTGGAAACAACCATTCATACAG -3' and 5'- CGATTAACTGCCTGGTGATTTTG -3'. For kinetic analysis at the one-cell stage, embryos were collected 12 hours after injection of Pibf1-

TALEN mRNA and processed using the GenomePlex Single Cell Whole Genome Amplification Kit according to the manufacturer’ s instructions (Sigma).

5. Western blot analysis

Proteins were extracted from tissue samples by homogenizing in a lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.5% deoxycholic acid, 0.1% SDS) supplemented with proteasome and phosphatase inhibitors, and western blot analysis of Sepw1 was performed with 30 μg of proteins using a Selenoprotein W-specific rabbit polyclonal antibody (Novus Biologicals, Littleton, Colorado, USA) according to manufacturer's instructions.

6. Preparation of TALEN mRNAs

Pibf1- and Sepw1-TALEN mRNAs were prepared using the mMESSAGE mMACHINE T7 Ultra kit (Ambion) according to the manufacturer’ s instructions and diluted to working concentrations in diethyl pyrocarbonate (DEPC, Sigma)-treated injection buffer (0.25 mM EDTA, 10 mM Tris, pH 7.4).

7. Microinjection of TALEN mRNAs into mouse embryos

All animal experiments were performed in accordance with the Korean Food and Drug Administration (KFDA) guidelines. Protocols were reviewed and approved by the Institutional Animal Care and Use Committees (IACUC) of the Laboratory Animal Research Center at Yonsei University (Permit Number: 2012-0087). All mice were maintained in the specific pathogen-free facility of the Laboratory Animal Research Center at Yonsei University.

C57BL/6 (B6) and ICR mouse strains were used as embryo donors and foster mothers, respectively. Female B6 mice (7-8 weeks old) were super-ovulated by intraperitoneal injections of 5 IU pregnant mare serum gonadotropin (PMSG, Sigma) and 5 IU human chorionic gonadotropin (hCG, Sigma) at 48-hour intervals. The super-ovulated female mice were mated to B6 stud males, and fertilized embryos were collected from oviducts.

TALEN mRNAs was injected into the cytoplasm of fertilized eggs with well recognized pronuclei in M2 medium (Sigma) using Piezo-driven micromanipulator (Prime Tech). For Pibf1-TALEN, two different concentrations (20 and 50 ng/ul of Pibf1-TALEN mRNA) were used, whereas a single dose of Sepw1-TALEN mRNA (50 ng/ul) was injected. After incubation at 37℃ for 24 hours, two-cell embryos were selected and transferred into the oviducts of pseudopregnant foster mothers.

8. Fluorescence PCR

Genomic Genomic DNA (100 ng per reaction) was subjected to PCR analysis using HiPi Plus DNA polymerase (ELPIS) and 5′ –carboxyfluorescein–labeled primers. The PCR amplicons were studied using an ABI 3730xl DNA analyzer.

Results

1. Activity test of TALENs in mouse cell line

To target these genes in the mouse genome, I designed and synthesized highly active TALENs specific to exon 2 of *Pibf1* (Pibf1-TALEN; Figure 1) and exon 1 of *Sepw1* (Sepw1-TALEN; Figure 2). To test the efficiency each TALENs, I transfected human embryonic kidney 293 (HEK-293) cells with these TALEN plasmids and a surrogate reporter (Kim et. al 2011). Flow-cytometric analyses measuring nonhomologous-end-joining (NHEJ) activity induced by Pibf1-TALEN and Sepw1-TALEN showed that about 13% and 21.5% of cells had frameshift indel mutations in the reporter plasmid, respectively (Figure 3).

I also transfected Pibf1-TALEN plasmids into NIH3T3 mouse cell line. T7E1 assay and sequencing analysis showed that these TALENs efficiently induced small indel mutations at the target site (Figure 3).

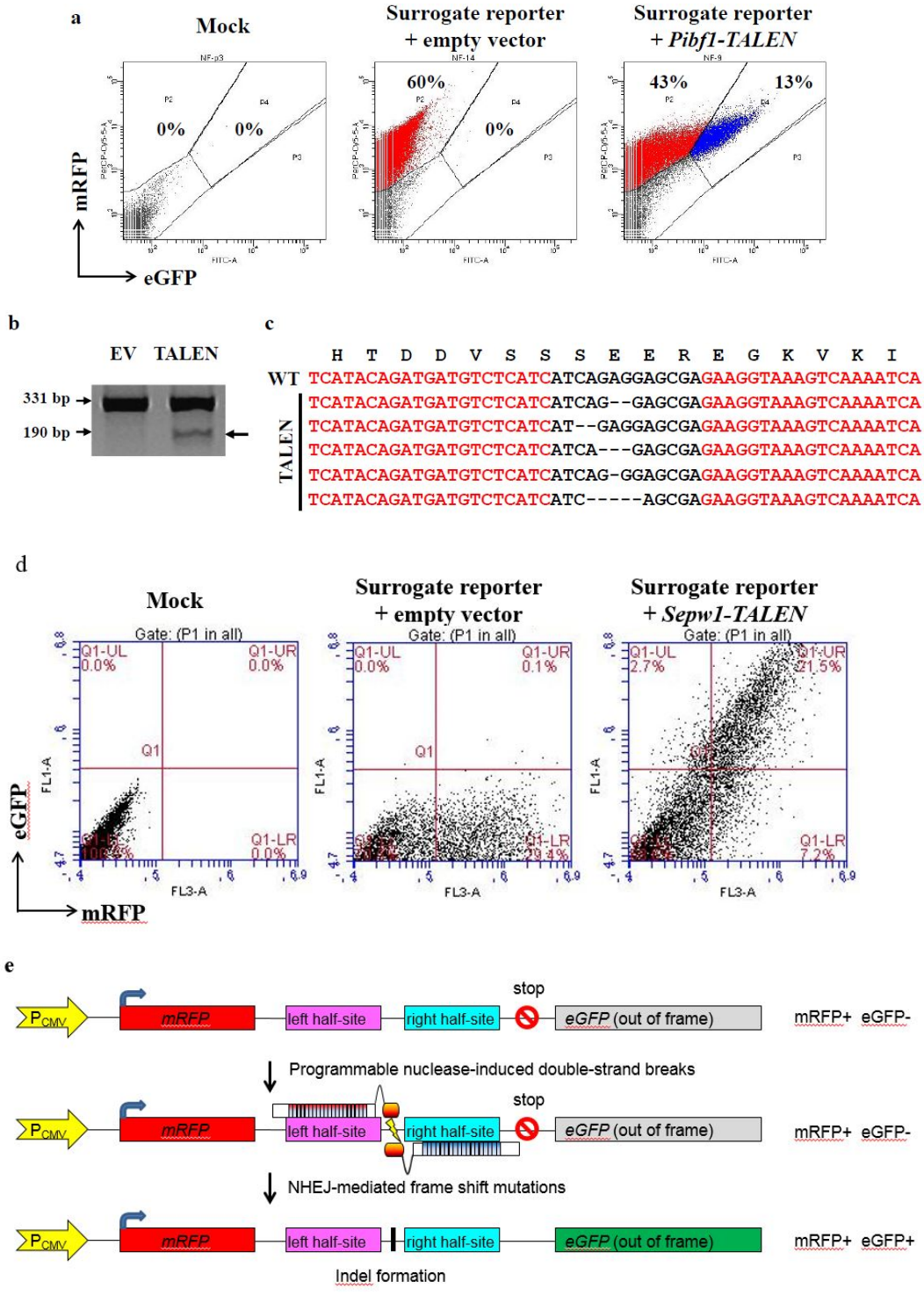

a
Mock
Surrogate reporter + empty vector
Surrogate reporter + Pibf1-TALEN
0%
0%
60%
0%
43%
13%
mRFP
eGFP
b
EV
TALEN
331 bp
190 bp
c
H T D D V S S S E E R E G K V K I
WT TCATACAGATGATGTCTCATCATCAGAGGAGCGAGAAGGTAAAGTCAAAATCA
TALEN
TCATACAGATGATGTCTCATCATCAG--GAGCGAGAAGGTAAAGTCAAAATCA
TCATACAGATGATGTCTCATCAT--GAGGAGCGAGAAGGTAAAGTCAAAATCA
TCATACAGATGATGTCTCATCATCA---GAGCGAGAAGGTAAAGTCAAAATCA
TCATACAGATGATGTCTCATCATCAG-GGAGCGAGAAGGTAAAGTCAAAATCA
TCATACAGATGATGTCTCATCATC-----AGCGAGAAGGTAAAGTCAAAATCA
d
Mock
Surrogate reporter + empty vector
Surrogate reporter + Sepw1-TALEN
Gate: (P1 in all)
Q1-UL 0.0%
Q1-UR 0.0%
Q1-LR 0.0%
Q1-UL 0.0%
Q1-UR 0.1%
Q1-LR 29.4%
Q1-UL 2.7%
Q1-UR 21.5%
Q1-LR 7.2%
Q1
FL1-A
FL3-A
eGFP
mRFP
e
stop
P CMV
mRFP
left half-site
right half-site
eGFP (out of frame)
mRFP+ eGFP-
Programmable nuclease-induced double-strand breaks
NHEJ-mediated frame shift mutations
mRFP+ eGFP+
Indel formation

Figure 3. The specific activity of Pibf1−TALEN. **(a)** Flow−cytometric analyses measuring nonhomologous−end−joining (NHEJ) activity induced by Pibf1−TALEN in human embryonic kidney 293 (HEK−293) cells. Populations in red (mRFP) indicated transfected cells, and GFP expression showed NHEJ activity. **(b)** T7E1 assays using NIH3T3 cells transfected with empty vector (EV) or Pibf1−TALEN (TALEN). The arrow indicates the product of T7E1 digestion. **(c)** Direct sequence analysis of the target region of the *Pibf1* locus of NIH3T3 cells, showing frameshift mutations induced by Pibf1−TALEN. Sequences in red indicate the TALEN binding sites. "−" denotes deleted nucleotides. **(d)** The specific activity of Sepw1−TALEN. Flow−cytometric analyses measuring NHEJ activity induced by Sepw1−TALEN in HEK−293 cells. Populations in red (mRFP) indicate transfected cells, and concomitant GFP expression reflects NHEJ activity. **(e)** Schematic overview of surrogate reporter system (Kim et al. 2011). When a double−strand break is introduced into the target sequence by programmable nucleases like TALENs, the break is repaired by nonhomologous end−joining (NHEJ), which often causes frameshift mutations. Such mutations can render eGFP in frame with mRFP, inducing the expression of the mRFP−eGFP fusion protein.

2. Generation of TALEN-mediated mutant mice

Each TALEN mRNA pair was injected into the cytoplasm of mouse pronuclear-stage embryos to produce mutant founders (F0) with mutations in *Pibf1* (Figure 4, 5, and Table 1) or *Sepw1* (Figure 6 and Table 2). Most TALEN-induced mutations were deletions of variable lengths that induced frameshifts in the *Pibf1* and *Sepw1* genes (Figure 4~6). In-frame mutations, as a result of deletions and substitutions of specific amino-acid residues, were also frequent (Figure 4). Such mutations will be beneficial for studying putative domain- or amino acid residue-specific functions of the gene products. Insertional mutations were only observed in two instances (Figure 5).

To investigate the dose-dependent effects of the TALEN mRNAs, two different concentrations of Pibf1-TALEN mRNA (50 ng/μl and 20 ng/μl) were used, yielding 29 mutants (55.8%) from 52 newborns (Figure 4, 5, and Table 1). Bi-allelic mutations were observed in seven mutant mice (Figure 4, 5). The mutation rate was approximately proportional to the injection dose of Pibf1-TALEN mRNA. Injection of a high concentration (50 ng/μl) of Pibf1-TALEN mRNA yielded 10/13 (76.9%) mutant F0 mice, whereas injection of a lower dose (20 ng/μl) yielded 19/39 (48.7%) F0 mutant mice (Table 1). Bi-allelic mutations were also found more frequently in the high-dose (6/8 F0) than in the low-dose (1/19 F0; Figure 4, 5, and Table 1) group, indicating that Pibf1-TALEN activity was dose dependent. Furthermore, relatively large deletions were

more frequently observed in mutant founders obtained by high-dose injection than in those by low-dose injection (Figure 4, 5).

In contrast to the mutation rate, the number of mutant mice produced by the low dose injection (19 mutants/176 transplanted embryos, 10.8%) was ~2.5 fold larger than the number produced by the high-dose injection (10 mutants/243 transplanted embryos, 4.1%; Table 1). This phenomenon is reminiscent of the toxicity of zinc-finger nucleases (ZFNs) induced by generation of nonspecific double-strand breaks at off-target sites (Szczepek et al. 2007). Indeed, IgM mutant rats obtained by intracytoplasmic injection of IgM-TALEN mRNA had modifications at an off-target site (Tesson et al. 2011). Although off-target effects of Pibf1-TALEN was not detected by T7E1 assays (Table 3 and Figure 7), it cannot be ruled out the possibility that nonspecific effects might have caused embryonic toxicity.

To validate TALEN-induced mutations, it is measured the level of Sepw1 protein in tail biopsies. No expression of Sepw1 protein was detected in a *Sepw1* mutant founder possessing bi-allelic null mutations at their start codon (Figure 8).

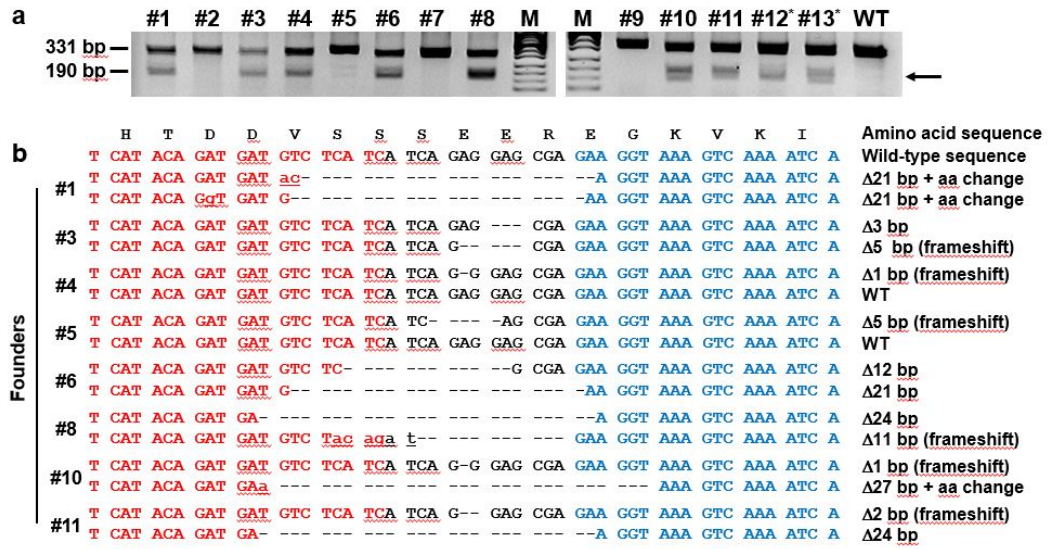
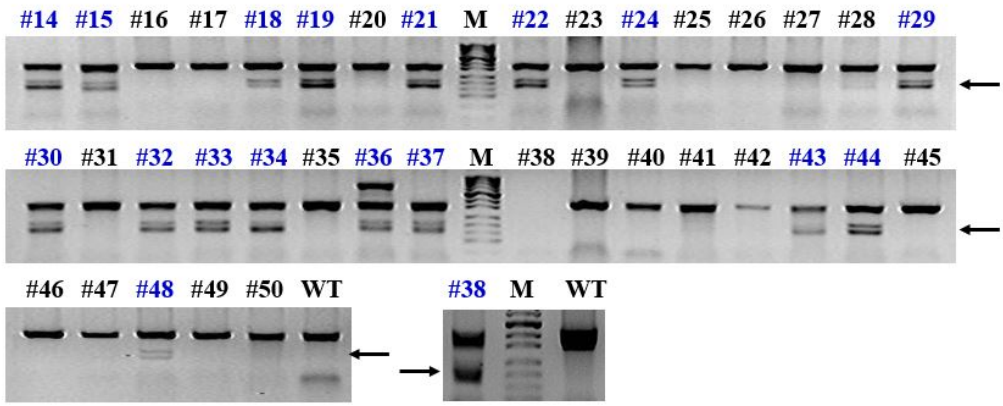


Figure 4. Generation of TALEN-mediated *Pibf1* mutant mice. **(a)** T7 endonuclease I (T7E1) assays were conducted using genomic DNA from pronuclear-stage embryos intra-cytoplasmically injected with Pibf1-TALEN mRNA (50 ng/μl). The arrow indicates the size of T7E1-digested DNA fragments. *Cannibalized mice. **(b)** DNA sequences of the *Pibf1* locus from live F0 mice identified by T7E1 assays in b. '-' denotes deleted nucleotides. Underlined sequences in lower case represent nucleotide substitutions.

a



b

Founder no.	Pattern 1	Pattern 2	Founder no.	Pattern 1	Pattern 2
14	WT	Δ2 bp	32	WT	Δ2 bp
15	WT	Δ32 bp	33	WT	Δ3 bp
18	WT	Δ11 bp	34	WT	Δ1 bp, Δ18 bp†
19	WT	Δ2 bp	36	WT	+321 bp††
21	WT	Δ2 bp	37	WT	Δ7 bp
22	WT	Δ2 bp	38	WT	Δ13 bp, Δ14 bp
24	WT	Δ3 bp	43	WT	Δ13 bp
28	Δ3 bp		44	WT	Δ2 bp
29	WT	Δ1 bp	48	WT	+3 bp
30	WT	Δ2 bp			

c

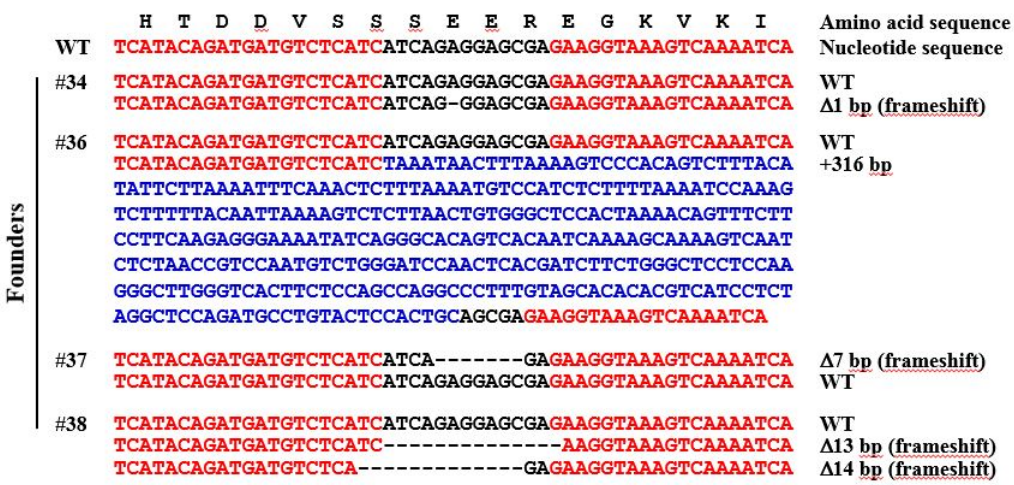


Figure 5. Generation of *Pibf1* mutant mice by intra-cytoplasmic injection of PN-stage embryos with a low dose (20 ng/ul) of Pibf1-TALEN mRNA. **(a)** T7E1 assays. The numbers in blue denote founder mice harboring mutations in the *Pibf1* locus. Arrows indicate the products of T7E1 digestion. **(b)** Fluorescence PCR (fPCR) analyses. †5bp longer than its sequence in c. Note that F0 mice #34 and #38 are genetically mosaic. **(c)** DNA sequences of the *Pibf1* locus in selected founder mice denoted in red in b. "-" denotes deleted nucleotides. *Note that Δ18-bp allele observed by fPCR in b was not identified.

Table 1. TALEN-mediated *Pibf1* targeting in C57BL/6J mice.

TALEN (ng/ml)	injected zygotes	surviving zygotes	Two-cell embryos	Transferred embryos	Newborns	Founders*
50	276	263(95.3%)	262(99.6%)	243	13(5.3%)†	10(76.9%)††
20	183	176(96.2%)	176(100%)	176	39(22.2%)†	19(48.7%)

Percentages were calculated using the number in each column as the numerator and the number in the column to its left as the denominator. *Determined by T7E1 assays. †Two pups were cannibalized at birth. ††T7E1 assays were conducted using genomic DNA samples from newborns, including the two cannibalized pups.

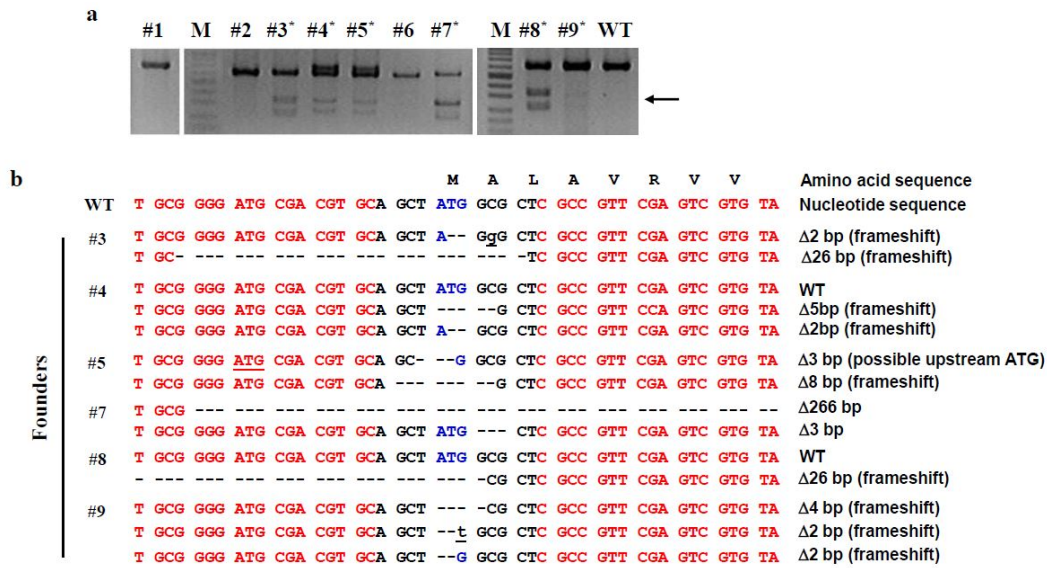


Figure 6. Generation of *Sepw1* mutant mice by intra-cytoplasmic injection of PN stage embryos with a high dose (50 ng/ul) of Sepw1-TALEN mRNA. **(a)** T7E1 assays for F0 screen. *, founder mice selected for direct sequence analyses shown in b. **(b)** DNA sequences of the *Sepw1* locus in founder mice. Sequences in red indicate TALEN binding sites; blue indicates the start codon (ATG). Underlined sequences in lowercase represent nucleotide substitutions. F0 mice #4 & #9 are genetically mosaic. Note that ⊿3 allele of F0 mouse #5 may have an upstream start codon (underlined ATG).

Table 2. TALEN-mediated *Sepw1* targeting in C57BL/6J mice.

TALEN (ng/ml)	Injected zygotes	Survived zygotes	Two-cell embryos	Transferred embryos	Newborns	Founders*
50	131	127(96.9%)	125(98.4%)	125	9(7.2%)†	6(66.7%)††

The percentages were calculated using the number in each column as the numerator and the number in the column to its left as the denominator. †Five pups were cannibalized at birth. ††T7E1 assays were conducted using genomic DNA samples from live pups. *Determined by T7E1 assays.

Table 3. Potential off-target sites of Pibf1-TALEN.

Locus	Sequences*	Pibf1-TALEN**
Pibf1	TCATACAGATGATGTCTCATC*atcagaggagcga*GAAGGTAAAGTCAAAATCA	TL1-TR2
Chr6: 63857551	TaATaaTGAaTcTAaaTTC*ctagtactttccag*GAtGGTAAAGTtAtgATCA	TR2-TR2
Chr1: 184373130	TGATTTTaACcTTACCcaA*tatagttcaaat*cAgGGTAAAaTCtAAATtA	TR2-TR2
Chr13: 94226483	TCAgACAGATcATGTCaacaC*taagaaatgagat*GAAGGTAAAGTCtActTCA	TL1-TR2
Chr9: 121769667	TaATaTTaAaaTTAtCTTC*attaaccaataat*GAAGaTAAAcctAAAATCA	TR2-TR2

* Putative binding sequences are in bold. Identical nucleotides are in uppercase.

** TL1, TL1-RR; TL2,TL2-DAS (see Materials and Methods)

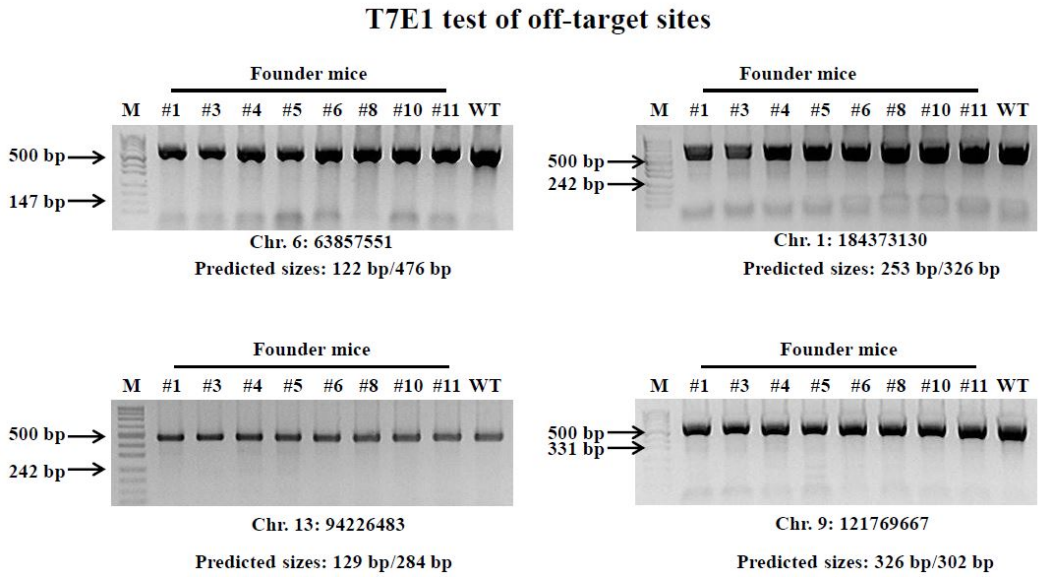
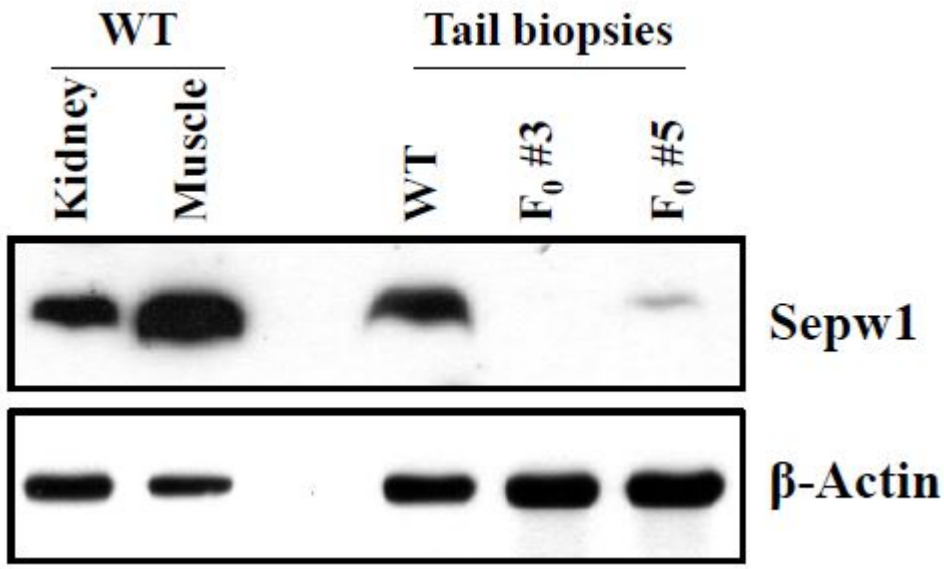


Figure 7. T7E1 assays examining the putative off-target effects of Pibf1-TALEN. The integrity of potential off-target sites listed in **Table 3** was examined in founder mice obtained by high-dose injection of Pibf1-TALEN mRNA.

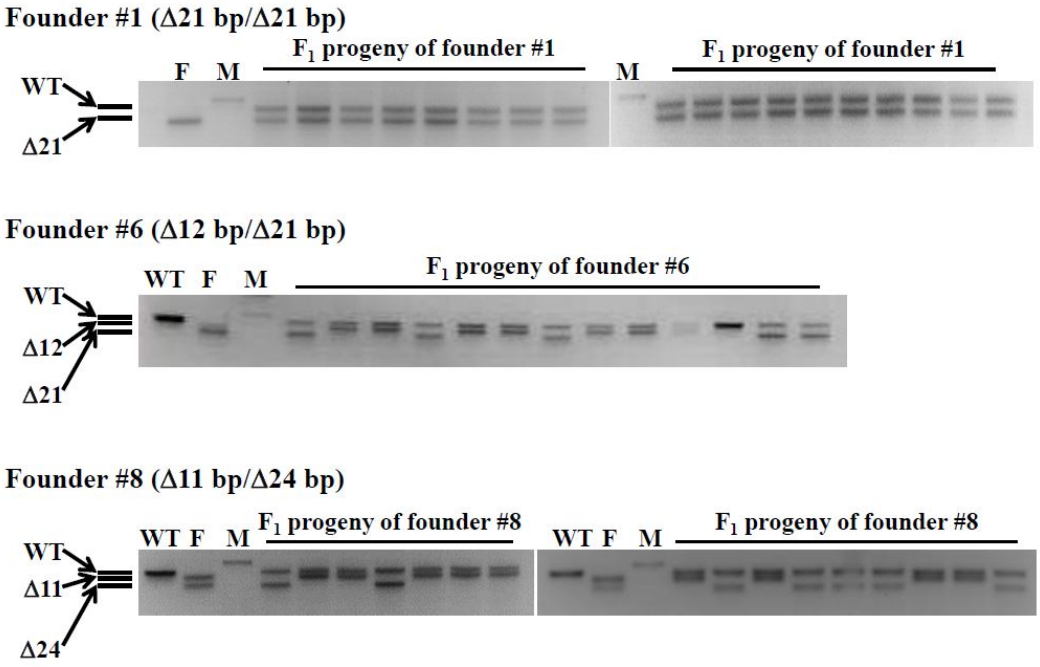


(by Lee K. in Yeungnam Univ.)

Figure 8. Expression level of Sepw1 protein in mutant mice. Western blot analysis of Sepw1 protein in tail biopsies of founders #3 & # 5 possessing bi-allelic null mutations in *Sepw1* locus. A very weak signal from F0 mouse #5 may indicate the inadequate translational product from ⊿3 allele indicated in **Figure 6**. Kidney as a tissue expressing a low level of Sepw1 protein; Muscle as a tissue showing a strong Sepw1 expression; WT, wild-type.

3. Germ line transmission of mutations

To confirm that mutations induced by TALENs were inherited to F1 progeny, three F0 *Pibf1* mutants were crossed to wild-type mice and genotypes of the F1 offspring were determined. All the mutations observed in F0 mice were transmitted through the germline (Figure 9). These results indicate that TALEN-introduced mutant alleles were stably inherited by their F1 progeny.



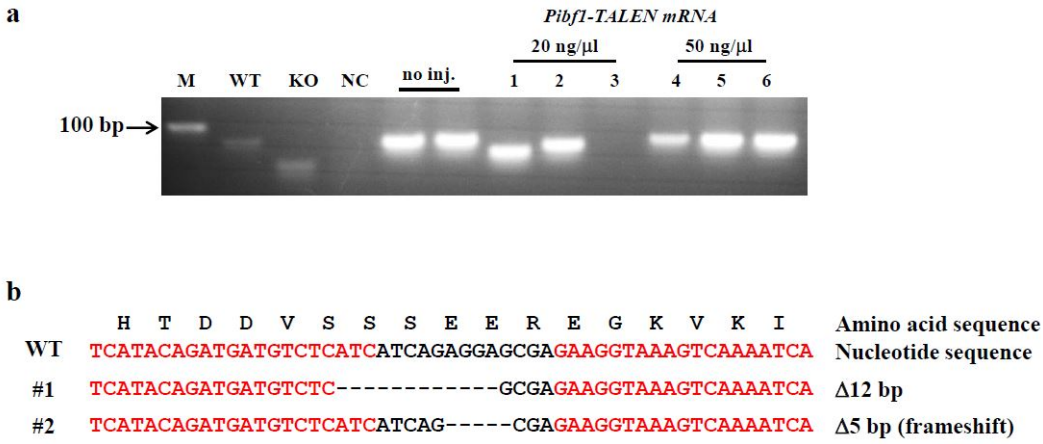
(by Sung YH. in Yonsei Univ.)

Figure 9. Germ-line transmission of TALEN-induced *Pibf1* mutant alleles. Founder mice harboring TALEN-induced mutations were crossed to wild-type mice, and the genotypes of their F1 progeny were determined by PCR genotyping. F, Founder mice (F0); WT, wild-type; M, 100-bp marker DNA.

4. Activity of TALENs at the one cell stage

Mosaicism was rarely detected in both *Pibf1* (Figure 4, 5) and *Sepw1* mutants (Figure 6), indicating that TALENs are primarily active at the one cell stage. To demonstrate that Pibf1-TALEN is active at this stage, the whole genomes of one-cell embryos was amplified before initiation of the first cell division and assessed the alteration of the target region in the *Pibf1* locus. PCR genotyping and sequencing revealed deletions in the *Pibf1* locus, even in one-cell embryos (Figure 10). I conclude that TALEN activity in one-cell embryos is sufficient to induce mutations.

The occurrence of mosaicism in F0 would be predicted to result from sustained TALEN activity during later embryogenesis or re-cleavage of already-modified alleles (Tesson et al. 2011). Because the first round of DNA replication begins after pronuclei become visibly evident under a stereomicroscope (Adenot et al. 1997), the presence of mosaic mutants would not necessarily prove that TALENs act at the two-cell stage or later. If TALEN activity were sustained or delayed until the late one-cell stage, when the embryo possesses 4N DNA content before cell division, up to four different mutant alleles would be produced. No more than three different alleles were found among the *Pibf1* (Figure 4, 5) or *Sepw1* mutant founders (Figure 6). These results support the hypothesis that TALEN-induced mosaic mutations can occur at the one-cell stage.



(by Sung YH. in Yonsei Univ.)

Figure 10. TALEN induces mutations in the *Pibf1* locus in one-cell embryos. Embryos were harvested 12 hours after intra-cytoplasmic injection of PN-stage embryos with Pibf1-TALEN mRNA. **(a)** PCR genotyping after whole genome amplification (WGA). WT, wild-type; KO, founder #1 (⊿21, bi-allelic mutant); NC, negative control (human genomic DNA); no inj., embryos that were not injected with Pibf1-TALEN mRNA. **(b)** DNA sequences of PCR products (#1 and #2 in **(a)**) exhibiting deletions in the *Pibf1* locus in DNA from one-cell embryos. Sequences in red indicate TALEN binding sites. "-" denotes deleted nucleotides.

Discussion

In this study, knockout of the *Pibf1* and *Sepw1* gene were successfully generated via transcription activator-like effector nucleases in NIH3T3 cell line and mice. Efficiency of yielding mutant mice is up to 76.9% depending on mRNA injection dose and bi-allelic mutations were also found very frequently without any off-target effects.

Efficient and precise genome engineering in living cells and organism hold great promise in both basic and applied research. These technologies exploit the ability of engineered nucleases to cause chromosomal DNA double-stranded breaks (DSBs) and stimulate the subsequent repair mechanisms like non-homologous end-joining (NHEJ) and homologous recombination (HR). NHEJ repair pathway results in insertion and deletion mutations in the targeted loci, whereas HR with donor DNA including homologous sequences of target region can generate replacement of target sequences. Although HR efficiency had not been tested in this study, recent papers report that HR via TALENs is very successful in many cells and organisms (Miller et al. 2011; Zu et al. 2013; Haibao Zhu et al. 2013; Katsuyama et al. 2013; Shin et al. 2014) including mice (Kasparek et al. 2014). Accordingly, TALEN technology has significant potential in experimental biology.

Zinc finger nuclease (ZFN) is also useful tool for genome engineering, but widespread adoption of ZFNs is hampered by some

problems. Zinc finger recognize 3-bp subsites and up to 64 different modules are required to assemble zinc finger array for recognizing all of 64 triplet bases. Unfortunately, unique zinc fingers targeting some 3-bp subsites are very hard to design (Maeder et al. 2008; Sander et al. 2011). Whereas, one TALE repeat module can recognize just single base pair and only four different modules are needed for making TALENs. Rearranging array of TAL modules, TALEN can recognize any wanted DNA sequences. Together with toxicity problem of ZFNs discussed above (Miller et al. 2007; Cornu et al. 2008; Radecke et al. 2010), TALEN is more convenient and useful method for genome engineering.

In this study, bi-allelic mutant of mice were frequently generated by TALENs (Figure 4~6) which means TALEN is active at the one cell stage. Although not in all tested embryos, activity of TALEN at this stage was detected even in low mRNA dose condition (Figure 10). If Pibf1-TALEN protein is still active after the first cleavage of one-cell embryos, it should mutate the target locus independently in each nucleus of the two blastomeres, thereby frequently generating mosaic animals. However, predominantly mono-allelic *Pibf1* mutations were produced in the low-dose experiment, indicating that the activity of Pibf1-TALEN was not sustained in blastomeres at the two-cell stage. This pattern is reminiscent of the maternal-to-zygotic transition (MZT) that actively eliminates maternally provided gene products (Tadros W et al. 2009). Although more detailed studies should be conducted that are designed to

provide more direct evidence for TALEN activity at different developmental stages, results of this study suggest that TALEN activity is not likely to be maintained after the first cleavage of one cell embryos which means it is very easy to generate knockout mice with TALENs.

In summary, this study establishes that TALEN mediated gene targeting is an efficient method for creating heritable null mutations in a specific locus of the mouse genome without any off-target effects. Furthermore, TALEN activity in one-cell embryos is sufficient to induce mutations. Taken together, these data suggest that TALEN-mediated in vivo mutagenesis might expedite the creation of genetically engineered mouse models and thereby help to accelerate functional genomic research. Thus, TALEN technology will be a powerful option that will facilitate precision genome engineering in bioresearch.

Chapter 2. Efficient gene correction of *F8* inversion in hemophilia A patients via CRISPR/Cas system

Introduction

Gene therapy is an experimental technique to correct defective genes in patient' s cells instead of using drugs or surgery. There are several gene therapeutic ways that: 1) gene addition approaches in which the therapeutic cDNA is expressed (Cappelli et al. 2010), 2) knocking out a mutated gene that is functioning improperly or receptor gene for viral infection (Li et al. 2013a) and 3) correct the mutant gene at the endogenous locus (Park et al. 2014). Gene targeting at the predetermined endogenous locus is very intriguing method because it can preserve not only target gene sequences but genomic context imparted by the native locus.

Engineered nucleases like ZFNs or TALENs are very useful tools for targeted genome editing. In recent studies, clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas) system is used to induce site-specific DSBs to modify the genome in a targeted manner (Cho et al. 2013a; Cong et al. 2013; Hwang et al. 2013; Friedland et al. 2013; Yang et al. 2014). Naturally, CRISPR/Cas system is an adaptive immune system to viruses and plasmids in bacteria or archaea (Wiedenheft et al. 2012). CRISPR RNA (crRNA), trans-activating RNA (tracrRNA) and Cas9 protein in type II CRISPR/Cas system make ribonucleoprotein (RNP) complex and performs as a sequence-specific endonuclease. Recent paper showed that the

crRNA works like a guide-RNA whose specificity is determined by spacer sequences (Jinek et al. 2012), so we can use this system for targeted genome engineering via designing of crRNA spacer sequences.

With the advent of molecular biology like DNA sequencing technology, diverse structural variations (SVs) of human genome were uncovered. SVs are generally defined as genomic alterations that involve segments of DNA larger than 1 Kb (Feuk et al. 2006) including deletion, insertion, duplication, translocation and inversion and these are associated with diverse genetic diseases (Stephens et al. 2009; Stankiewicz et al. 2010).

Hemophilia A (HA) is a representative genetic disease which is an X-chromosome-linked coagulation disorder with approximately 1 in every 5000 males worldwide (Graw et al. 2005). This disorder is caused by various genetic mutations in the X-linked coagulation *factor VIII* (*F8*) gene. HA can be characterized as severe (<1% activity), moderate (1-5% activity), and mild (5-30% activity), depending on the relative amount of F8 activity in the patient's plasma (Graw et al. 2005). Almost one half of patients with severe HA have large DNA inversions that disrupt either intron 1 (Inv1) or intron 22 (Inv22) of the *F8* gene (Bagnall et al. 2002; Naylor et al. 1993). Both Inv1 and Inv22 arise through non-allelic meiotic recombination between near-identical inversely oriented duplicons (i.e. *int1h* and *int22h*, respectively; Figure 11). Although six patients with hemophilia B, a less common form of X-linked bleeding disorder than

hemophilia A, treated with an adeno-associated virus vector (AAV) to deliver the *F9* cDNA which encodes blood coagulation *factor IX* (Nathwani et al. 2011), this vector cannot be used to deliver the full-length *F8* cDNA to patients with HA because AAV cannot accommodate the large size of the *F8* cDNA(~8 kbp). Thus, gene correction at endogenous *F8* locus is highly demanded for cure for severe hemophilia A.

Recently, Inv1 is successfully generated in human induced pluripotent stem cells (iPSC) with TALEN (Park et al. 2014). Introducing artificial inversion of about 140-kbp chromosomal segment between intron1 homologous sequences in human iPSC, hemophilia A model cell lines were created and successfully reverted to the wild-type state by reversing the segment. However, Inv1 is only about 4% of severe hemophilia A and artificial inversion of Inv22 has not been reported yet which is about half of severe HA. In addition, inversion with TALEN of this report is not tried in actual patient iPSC. Here, I have used a CRISPR/Cas system as a programmable RNA-guided engineered nucleases (RGENs) for introducing artificial inversion of Inv1 and Inv22 in Hela cells and patient iPSCs. Furthermore, mRNA expression of the *F8* gene is successfully restored in reverted patient iPSC clones.

Materials and Methods

1. Preparation of RGEN-encoding plasmids and recombinant Cas9 protein

Cas9-encoding plasmids were constructed as previously described (Cho et al. 2013a). Cas9 protein fused the HA epitope and a nuclear localization signal (NLS) was expressed under the control of the CMV promoter. U6 promoter was used to express a sgRNA as described previously (Cho et al. 2013a). Target sequences of each RGENs used in this study are shown in Table 5.

To purify Cas9 protein, the Cas9 DNA sequence was subcloned into pET28-b(+) (Invitrogen) vector. Expression of recombinant Cas9 protein containing a nuclear localization signal, the HA epitope, and the His-tag at the N-terminus was induced in BL21(DE3) strain using 0.5mM IPTG and cultured for 4hr at 25℃. The Cas9 protein was purified using Ni-NTA agarose beads (Qiagen), and dialyzed against 20mM HEPES pH 7.5, 150 mM KCl, 1mM DTT, and 10% glycerol. The purified Cas9 protein was concentrated using Ultracel 100K cellulose column (Millipore).

Guide RNA was in vitro transcribed using the MEGA shortscript T7 kit (Ambion) as previously described (Kim et al. 2014). Transcribed RNA was purified by phenol:chloroform extraction protocol, and purified

RNA was quantified using a spectrometer.

2. Cell cultures and transfections

HeLa (ATCC, CCL-2) was cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% antibiotics. To introduce DSBs, 1×10^5 HeLa cells were co-transfected with 0.5 ug of Cas9-encoding plasmid and sgRNA-encoding plasmid, respectively, using a Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. Human ESC (hESC) lines (H9) obtained from WiCell Inc., human dermal fibroblast-derived wild-type iPSCs (WT-iPSC Epi3 line) (Park et al. 2014), and urine-derived iPSCs generated in this study were maintained in hESC medium [DMEM/F12 (Gibco) supplemented with 20% knockout serum replacement (Gibco), 1% nonessential amino acids (Invitrogen), and 0.1 mM 2-mercaptoethanol (Sigma)] including 4 ng/mL basic FGF (PeproTech) as previously described (Kim et al. 2010; Jang et al. 2012). To introduce the inversion events in urine-derived iPSCs, 1×10^6 cells were electroporated with 5 ug of Cas9-encoding plasmid and sgRNA-encoding plasmid (5 ug of each sgRNA plasmid for intron 22 inversion), respectively, using a microporator system (Neon; Invitrogen). For direct delivery of Cas9 protein into the urine-derived iPSCs, transfection was performed as previously described (Kim S. et al. 2014) with slight modifications. Cas9 protein (15 ug) was mixed with 20 ug of transcribed sgRNA (20 ug of each sgRNA plasmid for intron 22 inversion) and incubated to formation

of RGEN ribonucleoprotein complexes (RNPs) for 10 min at room temperature. The RNPs were transfected into 2×10^5 iPS cells by using a microporator system.

3. T7E1 assay and determining the frequencies of targeted inversion

The T7E1 assays were performed as previously described (Kim et al. 2009). Briefly, the genomic region encompassing the RGEN target site was PCR-amplified, melted, and re-annealed to form heteroduplex DNA, which was treated with 5 units of T7 endonuclease 1 (New England Biolabs) for 15 min at 37° C and then analyzed by agarose gel electrophoresis.

The frequencies of targeted corrections in the *F8* locus were estimated by digital PCR analysis as previously described (Lee et al. 2010). Briefly, the genomic DNA isolated from cell co-transfected with RGEN and sgRNA plasmids using lipofectamine 2000 (Invitrogen) was serially diluted, and the diluted samples were subjected to PCR analysis. The fraction of positive bands at each dilution points was counted and results were analyzed using the Extreme Limiting Dilution Analysis program (Hu et al. 2009).

4. Validation of RGEN mediated inversion of the *F8* locus in HeLa cells

To validate the genome editing activities of the RGENs designed for this study, each RGENs were co-transfected with sgRNA expression plasmid into HeLa cells and their activities were measured using the T7E1 assay and digital PCR analysis as described above.

5. Isolation and expansion of urine-derived cells from severe hemophilia A patients

The urine samples were collected from 11 severe patients who have been diagnosed with clinically confirmed by Korea Hemophilia Foundation Clinic. The urine-derived cells from patients were isolated as described previously (Zhou et al. 2012). In brief, the cells were collected by centrifuged at 400g for 10 min from approximately 100 ml of midstream urine sample. After washing twice with PBS, the cells were cultured in DMEM/Ham's F12 (1:1) medium (Hyclone) supplemented with 10% (vol/vol) fetal bovine serum (FBS), renal epithelial cell growth medium (REGM) SingleQuot kit (Lonza), and 1% antibiotics. Four day after culturing, the cells were subsequently cultured in renal epithelial basal medium (Lonza) supplemented with REGM SingleQuot kit for expansion of urine-derived cells. For further expansion, the cells were splited onto gelatin-coated culture dish at 80-90% confluency. To

confirm of *F8* genotype, genomic DNAs isolated from urine-derived patients cells were subjected PCR analysis with specific primer sets (Table 4) against intron 1 and 22 inversion of the *F8* locus as described previously (Park et al. 2014; Bagnall et al. 2006).

Table 4. Primer pairs used in this study.

Primer	Sequence (5' to 3')	Used for the experiment of
GAPDH-F	CCCCTCAAGGGCATCCTGGGCTA	qPCR / RT-PCR
GAPDH-R	GAGGTCCACCACCCTGTTGCTGTA	qPCR / RT-PCR
Oct4-F	CCTCACTTCACTGCACTGTA	qPCR
Oct4-R	CAGGTTTTCTTTCCCTAGCT	qPCR
Lin28-F	AGCCATATGGTAGCCTCATGTCCGC	qPCR
Lin28-R	TCAATTCTGTGCCTCCGGGAGCAGGGTAGG	qPCR
Sox2-F	TTCACATGTCCCAGCACTACCAGA	qPCR
Sox2-R	TCACATGTGTGAGAGGGGCAGTGTGC	qPCR
Nanog-F	TGAACCTCAGCTACAAACAG	qPCR
Nanog-R	TGGTGGTAGGAAGAGTAAAG	qPCR
F8-exon1-F	CTGCTTTAGTGCCACCAGAAGA	RT-PCR
F8-exon3-R	GACTGACAGGATGGGAAGCC	RT-PCR
F8-exon21-F	CCGGATCAATCAATGCCTGGAG	RT-PCR
F8-exon23-R	ATGAGTTGGGTGCAAACGGATG	RT-PCR
Brachyury-F	ATCACAAAGAGATGATGGAGGAA	RT-PCR
Brachyury-R	GGTGAGTTGTCAGAATAGGTTGG	RT-PCR
GAPDH-F	GAACATCATCCCTGCCTCTACTG	iPS generation (PCR)
GAPDH-R	CAGGAAATGAGCTTGACAAAGTGG	iPS generation (PCR)
EBNA-1-F	ATGGACGAGGACGGGGAAGA	iPS generation (PCR)
EBNA-1-R	GCCAATGCAACTTGGACGTT	iPS generation (PCR)

(continued)

Primer name	Sequence (5’ to 3’)	Used for the experiment of
1-H1	CTCCAGCCACCAGAACCA	T7E1 assay
1-H2	GCCAGTCTCTCCTTGTGTGT	T7E1 assay
1-F1	AAATCACCCAAGGAAGCACA	Genotype PCR
1-R1	TGGCATTAACGTATTACTTGGAGA	Genotype PCR
1-F2	TGCTGAGCTAGCAGGTTTAATG	Genotype PCR
1-R2	TGCTGAGCTAGCAGGTTTAATG	Genotype PCR
22-F1	TGGGGCTGTGTAAATTTGCT	Genotype PCR
22-R2	CAAACGTAGCATTACCTGATTGT	Genotype PCR
22-F2	ACAACCAGAGCAGAAATCAATGA	Genotype PCR
22-R2	TTTCACCACATCCACGCCAA	Genotype PCR

6. Targeted corrections of the *F8* locus in patient derived-iPSCs via RGEN plasmids

The iPSCs were cultured onto STO feeder layer and harvested by treating with collagenase type IV. After washing with PBS, the cells were dissociated into approximately single cell level as previously described (Desbordes et al. 2008). These single cells were mixed with RGEN and sgRNA plasmids and pulsed with a voltage of 850 for 30 ms. Cells were then seeded onto feeder cells and allowed to grow for 10 days. To detect genomic inversion events occurring in F8 locus, cells from individual colonies were lysed and subjected to PCR as described previously (Park et al. 2014). PCR products were analyzed by agarose gel electrophoresis. Specific primer sequences are shown in Table 4.

7. Isolation of clonal populations of cells, PCR analysis, and DNA sequencing of breakpoints

To isolate clonal populations of corrected cells, each colony that had been identified by PCR as containing the desired genomic event (namely, correction of inversion genotype) was dissociated into single cells and re-seeded onto new feeder layer. After 4 rounds of passaging, several clones (4 clones for intron 1 correction, 3 clones for intron 22 correction) were chosen for sequencing and further experiments. To determining the sequence at the breakpoints, amplified PCR products were

electrophoresed, and PCR products were eluted from the agarose gel as previously described (Park et al. 2014).

8. RNA isolation, RT-PCR, and qPCR

Total RNAs were purified from cells using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. cDNAs were synthesized from total RNAs (1ug) using the DiaStarTM cDNA synthesis kit (SolGent, Korea). To confirm the expression of Factor VIII, Brachyury, and GAPDH, PCR was performed with Ex-Taq (Takara) using the synthesized cDNAs as template. For qPCR, SYBR® Premix Ex-Taq (Takara) was used according to the manufacturer's instructions. To amplification mRNA of *F8* from intron 1 or intron 22 corrected lines, respectively, a forward primer located in exon 1 (or exon 21 for intron 1 corrected lines) was used in combination with a reverse primer located in exon 3 (or exon 23 for intron 22 corrected lines). Specific primer sequences used for RT-PCR or qPCR are shown in Table 4.

9. Generation of iPSCs from urine-derived cells and *in vitro* differentiation

Less than 3 time passage, the urine-derived cells were used for iPS cell (iPSC) generation. Episomal reprogramming vectors were used as previously reported (Okita et al. 2011) for iPSC generation from two

urine-derived patients cells (referred as Pa1-UC and Pa2-UC) among three patients cells. Sendai virus purchased from Invitrogen was used for iPSC generation from third urine-derived patient cells (referred as Pa3-UC) according to the manufacturer's instructions. Seven iPSC colonies that looked similar to human ES cells were picked up mechanically, and were further cultured for characterization. In vitro differentiation of the iPSCs into three germ layers was performed as previously described (Sugii et al. 2010; Jang et al. 2011). For induction of embryoid bodies (EBs), iPS colonies on feeder cells were partially dissociated using collagenase type IV (Invitrogen). EBs were transferred to Petri dishes (SPL Lifesciences, Korea) and cultured in hESC medium without basic fibroblast growth factor (bFGF) but supplemented with 5% FBS for ten days. Spontaneous differentiation of EBs into cells representing the three germ layer lineages was detected by immunostaining with appropriate antibodies. To induce differentiation of iPSCs into the mesoderm lineage, we used a previously described report (Yoo et al. 2013) with slight modifications. Briefly, EBs were transferred to Petri dishes and cultured in hESC medium excluding basic fibroblast growth factor (bFGF) but supplemented with 20 ng/mL bone morphogenic protein-4 (BMP4, R&D Systems) and 10 ng/mL Activin A (PeproTech). On day 3, EBs were attached onto Matrigel-coated dishes and were cultured for an additional 3 days in same medium described above. On day 6 of differentiation, cells that had differentiated into the mesoderm lineage were harvested for

determination of *F8* gene expression.

10. Characterization of iPSCs

Alkaline phosphatase (AP) activity was measured with the leukocyte alkaline phosphatase staining kit (Sigma) according to the manufacturer's instructions. To confirm the urine-derived cells origin of iPSC lines, short tandem repeat (STR) was analysed. STR loci were amplified from genomic DNA samples isolated from iPSC lines and their parental cells using the AmpFISTR PCR reaction system (Applied Biosystems). PCR-based STR analysis was peformed at Human Pass Inc. (Korea). For karyotype analysis, A G-banding analysis of chromosomes from the each iPSCs was performed at the GenDix Inc. (Korea). Immunostaining of ES cell markers was carried out as described previously (Park et al. 2014). DAPI (4', 6-Diamidino-2-Phenylindole, Vector Laboratories) was used for nuclei visualization. The images were captured and analyzed using an Olympus IX71 microscope or FSX system.

11. Analysis of off-target effects

Potential off-target sites of RGENs used in this study were searched *in silico* (Bae et al. 2014). RGEN-encoding plasmids were transfected in HeLa cells. Several potential off-target sites that similar to the RGEN target site were selected and amplified by PCR. The T7E1

analysis was performed as described above.

12. Ethical statement

Yonsei University Institutional Review Board approval (IRB # 4-2012-0028) was obtained for the generation and analyses of urine-derived iPS cells from the hemophilia A patients. All volunteers who participated in this study have signed written informed consent to donate urine sample for human iPSCs generation.

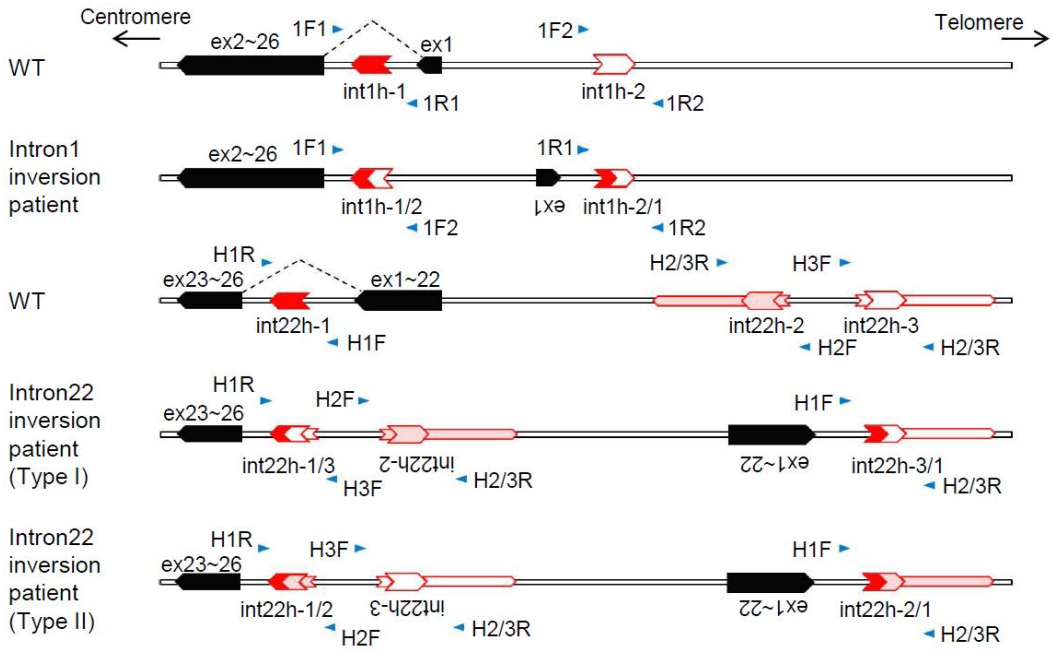


Figure 11. Two inversions in the *F8* gene. Schematic view of proposed model against the intron 1 and 22 inversions found in patients with severe hemophilia A. PCR primers used to detect each inversions are shown.

Results

1. Targeted inversion of intron 1 of *F8* via RGEN in Hela cells

First, I designed a RGEN for targeted reversion of intron1 in the *F8* gene (RGEN01) which recognize 20bp sequences in the homologous region of intron1. There are two homologous regions of intron1 in the human genome (here after int1h-1 and int1h-2); one is in the intron1 of the *F8* gene, and another is in the intergenic region located approximately 140-kb from the *F8* gene. Thus, this RGEN cut two regions together and inversion of 140kb genomic fragment between these sites could be occurred (Figure 12; Park et al. 2014).

To test the activity of this RGEN01, I transfected Cas9 plasmid and crRNA plasmid into Hela cells. I found that RGEN01 generated mutations most actively with an incidence of 34% at the target site (Figure 13A). In addition, the inversion specific bands between two homologous regions were detected in cells transfected with RGEN 01 (Figure 13B). Small deletion and insertions were found in breakpoint junction by sequence analysis from these PCR products (Figure 13 C, D). This 140-kb inversion that involves the intron 1 homolog was generated with a frequency of 3.1 % (for intron 1 homolog 1) and 2.2 % (for intron 1 homolog 2) in Hela cell (Figure 14).

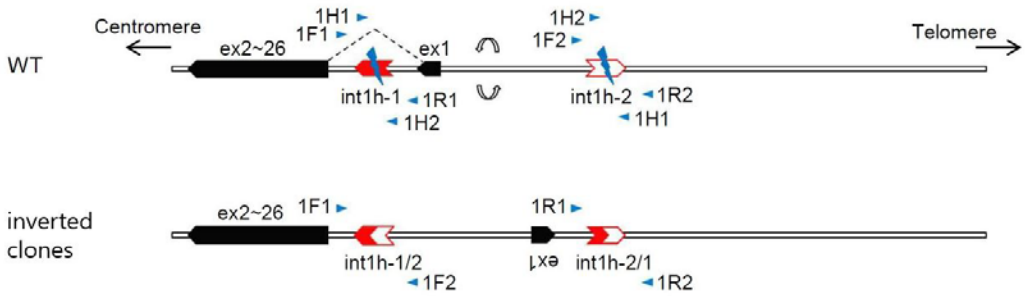


Figure 12. Schematic view of targeted inversion of intron1 in the *F8* gene in Hela cells. Two RGEN target sites recognized by one crRNA in the homologous region and PCR primers used to detect inversion are shown.

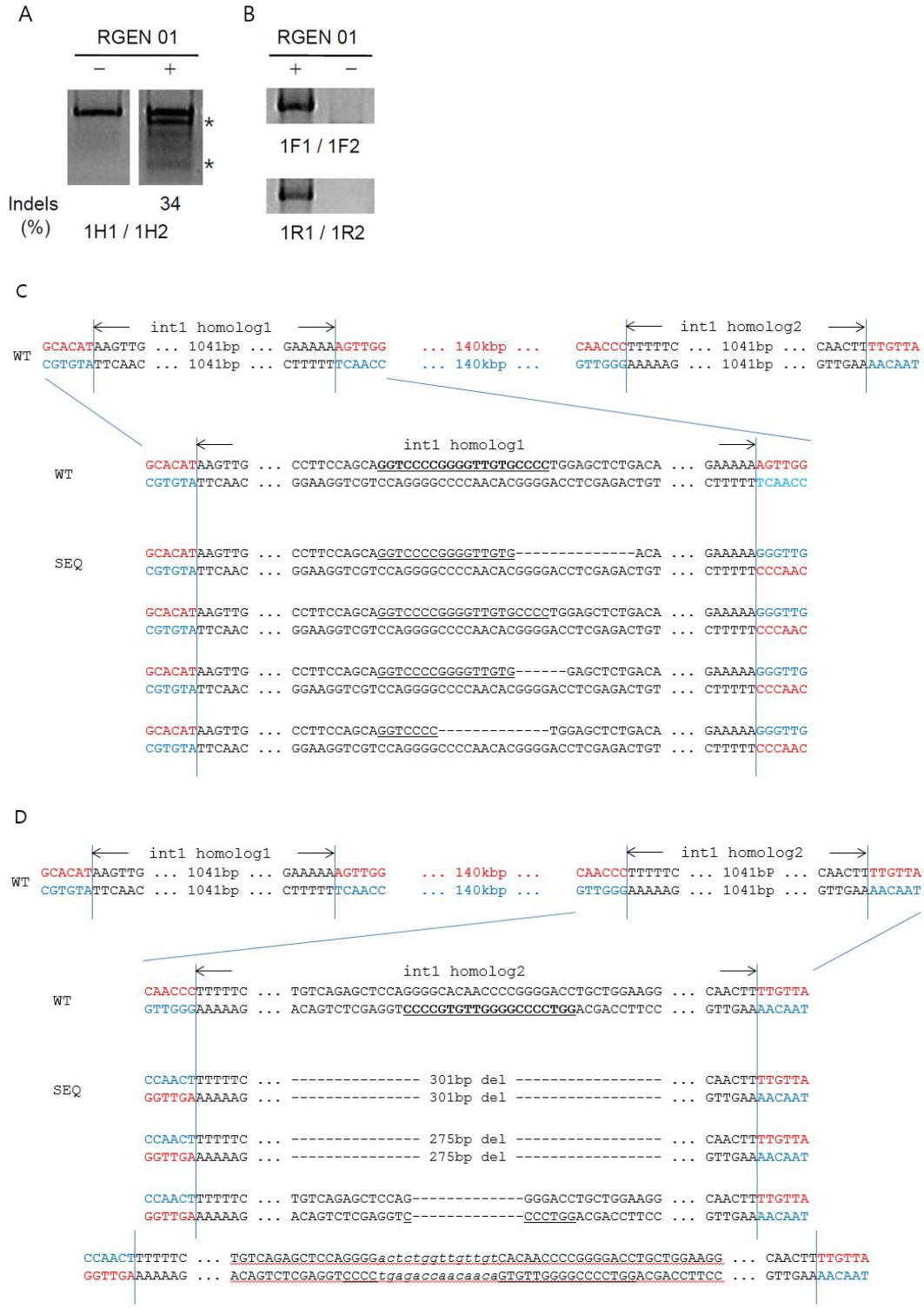
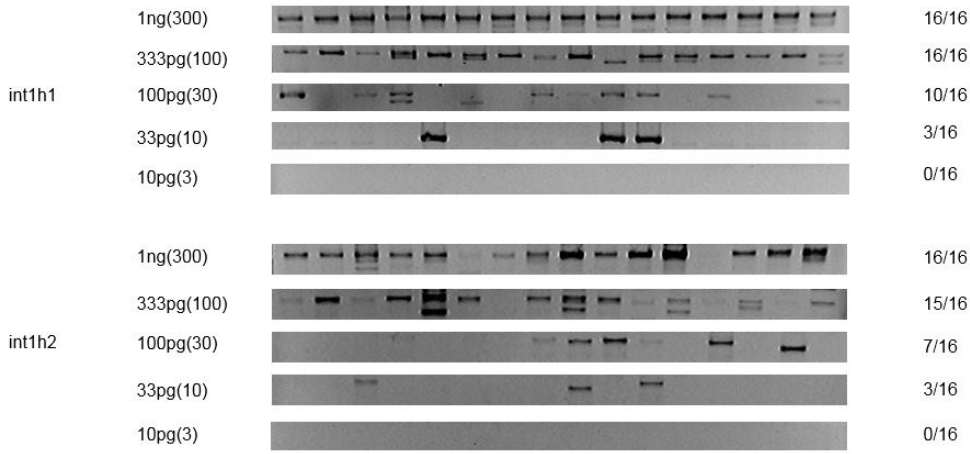

A
RGEN 01
−
+
*
*
Indels (%)
34
1H1 / 1H2
B
RGEN 01
+
−
1F1 / 1F2
1R1 / 1R2
C
int1 homolog1
int1 homolog2
WT
GCACATAAGTTG ... 1041bp ... GAAAAAAGTTGG ... 140kbp ... CAACCCTTTTTC ... 1041bp ... CAACTTTTGTTA
CGTGTATTCAAC ... 1041bp ... CTTTTTTCAACC ... 140kbp ... GTTGGGAAAAAG ... 1041bp ... GTTGAAAACAAT
int1 homolog1
WT
GCACATAAGTTG ... CCTTCCAGCAGGTCCCCGGGGTTGTGCCCCTGGAGCTCTGACA ... GAAAAAAGTTGG
CGTGTATTCAAC ... GGAAGGTCGTCCAGGGGCCCCAACACGGGGACCTCGAGACTGT ... CTTTTTTCAACC
SEQ
GCACATAAGTTG ... CCTTCCAGCAGGTCCCCGGGGTTGTG--------------ACA ... GAAAAAGGGTTG
CGTGTATTCAAC ... GGAAGGTCGTCCAGGGGCCCCAACACGGGGACCTCGAGACTGT ... CTTTTTCCCAAC
GCACATAAGTTG ... CCTTCCAGCAGGTCCCCGGGGTTGTGCCCCTGGAGCTCTGACA ... GAAAAAGGGTTG
CGTGTATTCAAC ... GGAAGGTCGTCCAGGGGCCCCAACACGGGGACCTCGAGACTGT ... CTTTTTCCCAAC
GCACATAAGTTG ... CCTTCCAGCAGGTCCCCGGGGTTGTG------GAGCTCTGACA ... GAAAAAGGGTTG
CGTGTATTCAAC ... GGAAGGTCGTCCAGGGGCCCCAACACGGGGACCTCGAGACTGT ... CTTTTTCCCAAC
GCACATAAGTTG ... CCTTCCAGCAGGTCCCC-------------TGGAGCTCTGACA ... GAAAAAGGGTTG
CGTGTATTCAAC ... GGAAGGTCGTCCAGGGGCCCCAACACGGGGACCTCGAGACTGT ... CTTTTTCCCAAC
D
int1 homolog1
int1 homolog2
WT
GCACATAAGTTG ... 1041bp ... GAAAAAAGTTGG ... 140kbp ... CAACCCTTTTTC ... 1041bP ... CAACTTTTGTTA
CGTGTATTCAAC ... 1041bp ... CTTTTTTCAACC ... 140kbp ... GTTGGGAAAAAG ... 1041bp ... GTTGAAAACAAT
int1 homolog2
WT
CAACCCTTTTTC ... TGTCAGAGCTCCAGGGGCACAACCCCGGGGACCTGCTGGAAGG ... CAACTTTTGTTA
GTTGGGAAAAAG ... ACAGTCTCGAGGTCCCCGTGTTGGGGCCCCTGGACGACCTTCC ... GTTGAAAACAAT
SEQ
CCAACTTTTTTC ... --------------- 301bp del ----------------- ... CAACTTTTGTTA
GGTTGAAAAAG ... --------------- 301bp del ----------------- ... GTTGAAAACAAT
CCAACTTTTTTC ... --------------- 275bp del ----------------- ... CAACTTTTGTTA
GGTTGAAAAAG ... --------------- 275bp del ----------------- ... GTTGAAAACAAT
CCAACTTTTTTC ... TGTCAGAGCTCCAG-------------GGGACCTGCTGGAAGG ... CAACTTTTGTTA
GGTTGAAAAAG ... ACAGTCTCGAGGTC-------------CCCTGGACGACCTTCC ... GTTGAAAACAAT
CCAACTTTTTTC ... TGTCAGAGCTCCAGGGGactctggttgttgtCACAACCCCGGGGACCTGCTGGAAGG ... CAACTTTTGTTA
GGTTGAAAAAG ... ACAGTCTCGAGGTCCCCtgagaccaacaacaGTGTTGGGGCCCCTGGACGACCTTCC ... GTTGAAAACAAT

Figure 13. Targeted correction of *F8* intron 1 locus in Hela cells. **(a)** Mutations at the RGEN 01 target site within intron 1 of *F8* were estimated in Hela cells by the T7E1 assay. Asterisks indicate the predicted position of DNA bands cleaved by T7E1. **(b)** PCR products corresponding to inversion genotypes. **(c, d)** DNA sequences of two homolog (here named int1 homolog 1 and 2) and breakpoint junctions in RGEN01 transfected cells. The RGEN target sequence is underlined.



	Amount of genomic DNA(Copy number per half genome)					Frequency(%)		p value (Fit)
	1ng	333pg	100pg	33pg	10pg		Upper and	
	(300)	(100)	(30)	(10)	(3)	Estimate	Lower limits	
int1h1	16/16	16/16	10/16	3/16	0/16	3.1	(2.0~4.7)	0.04
int1h2	16/16	15/16	7/16	3/16	0/16	2.2	(1.4~3.3)	0.24

Figure 14. Frequencies of targeted intron1 inversion. The digital PCR was performed for estimating the frequency of targeted inversion. Genomic DNA was isolated from the HeLa cells transfected with RGEN01 and was serially diluted. Estimated frequency of targeted inversions events created via RGEN01 was measured by digital PCR analysis (Lee H. et al. 2010). Upper and Lower limits indicate 95% confidence intervals.

2. Targeted inversion of intron22 of *F8* via RGEN in Hela cells

Intron 22 inversions is occurred by homologous recombination between three intron 22 homologous copies (refer to int22h-1, int22h-2, and int22h-3, respectively) (Lakich et al. 1993; Bagnall et al. 2006). The intron 22 homolog 1 (int22h-1) region is located in the *F8* gene locus, and others located approximately 500-kb (int22h-2) and 600-kb (int22h-3) apart from int22h-1, respectively (Figure 15). Unlike RGEN01 cutting two homologous regions of intron 1, RGEN that targets within int22h is able to cut three homologous copies, respectively. Then, two genomic fragments can be generated and variable unwanted (i.e. only one of them is inverted) clones could be more appeared. To solve this problem, I generated two specific RGENs (termed RGEN 02 and RGEN 03) that target nearby of int22h-1 or int22h-3 (Figure 15), and then tested their activity for genome engineering in Hela cell. RGEN02 and RGEN03 created indel mutations with an incidence of 44% and 32% at the each target sites, respectively (Figure 16A).

Next, I tested whether 600-kb genomic inversion between int22h-1 and int22h-3 could be generated in Hela cells containing wild-type genome structure. As shown in Figure 16B, this inversion was detected in cells that transfected with both RGENs. In each breakpoint junction, indels and large deletions were found by sequence analysis from

these PCR products (Figure 16C). This event was induced with a frequency of 2.2% (for breakpoint junction nearby intron 22 homolog 1) and 1.5% (for breakpoint junction nearby intron 22 homolog 3) in Hela cells by using digital PCR analysis (Figure 17).

Figure 15. Schematic view of targeted inversion of intron22 in the *F8* gene in Hela cells. Two target sites of each RGENs and PCR primers used to detect inversion are shown.

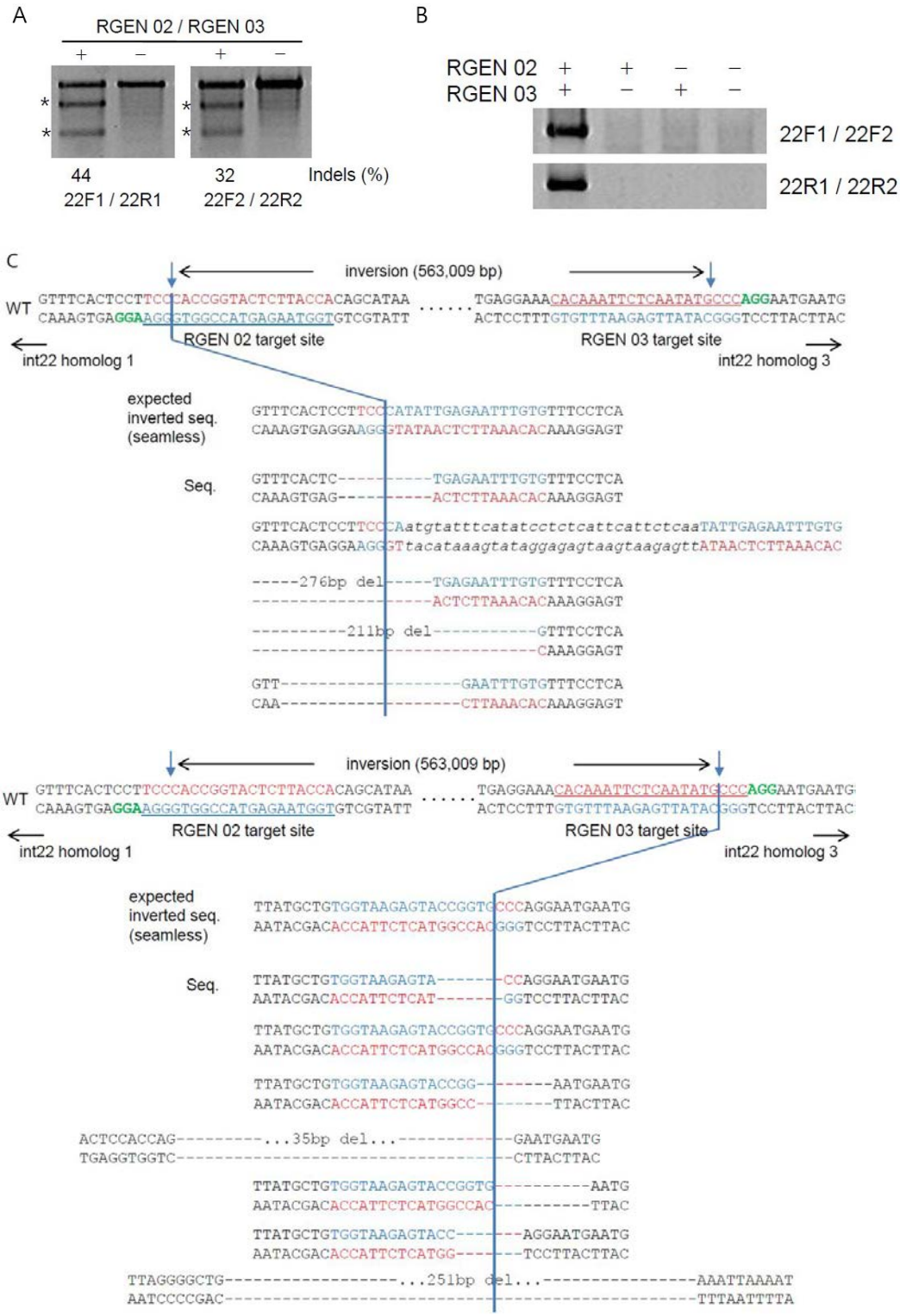
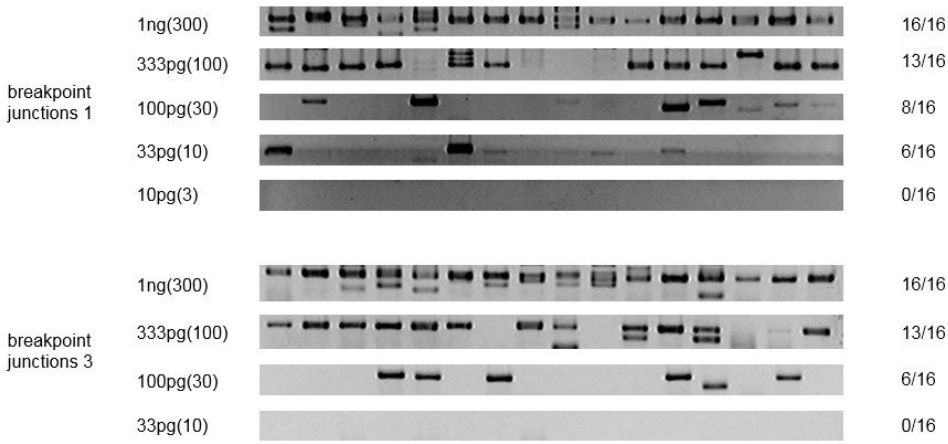

A
RGEN 02 / RGEN 03
+ − + −
*
*
44
22F1 / 22R1
32
22F2 / 22R2
Indels (%)
B
RGEN 02 + + − −
RGEN 03 + − + −
22F1 / 22F2
22R1 / 22R2
C
inversion (563,009 bp)
WT GTTTCACTCCTTCCCACCGGTACTCTTACCACAGCATAA ..... TGAGGAAACACAAATTCTCAATATGCCCAGGAATGAATG
CAAAGTGAGGAAGGGTGGCCATGAGAATGGTGTCGTATT ..... ACTCCTTTGTGTTTAAGAGTTATACGGGTCCTTACTTAC
RGEN 02 target site
RGEN 03 target site
int22 homolog 1
int22 homolog 3
expected inverted seq. (seamless)
GTTTCACTCCTTCCCATATTGAGAATTTGTGTTTCCTCA
CAAAGTGAGGAAGGGTATAACTCTTAAACACAAAGGAGT
Seq.
GTTTCACTC----------TGAGAATTTGTGTTTCCTCA
CAAAGTGAG----------ACTCTTAAACACAAAGGAGT
GTTTCACTCCTTCCCAatgtatttcatatcctctcattcattctcaaTATTGAGAATTTGTG
CAAAGTGAGGAAGGGTtacataaagtataggagagtaagtaagagttATAACTCTTAAACAC
-----276bp del-----TGAGAATTTGTGTTTCCTCA
----------ACTCTTAAACACAAAGGAGT
----------211bp del----------GTTTCCTCA
----------CAAAGGAGT
GTT----------GAATTTGTGTTTCCTCA
CAA----------CTTAAACACAAAGGAGT
inversion (563,009 bp)
WT GTTTCACTCCTTCCCACCGGTACTCTTACCACAGCATAA ..... TGAGGAAACACAAATTCTCAATATGCCCAGGAATGAATG
CAAAGTGAGGAAGGGTGGCCATGAGAATGGTGTCGTATT ..... ACTCCTTTGTGTTTAAGAGTTATACGGGTCCTTACTTAC
RGEN 02 target site
RGEN 03 target site
int22 homolog 1
int22 homolog 3
expected inverted seq. (seamless)
TTATGCTGTGGTAAGAGTACCGGTGCCCAGGAATGAATG
AATACGACACCATTCTCATGGCCACGGGTCCTTACTTAC
Seq.
TTATGCTGTGGTAAGAGTA-------CCAGGAATGAATG
AATACGACACCATTCTCAT-------GGTCCTTACTTAC
TTATGCTGTGGTAAGAGTACCGGTGCCCAGGAATGAATG
AATACGACACCATTCTCATGGCCACGGGTCCTTACTTAC
TTATGCTGTGGTAAGAGTACCGG--------AATGAATG
AATACGACACCATTCTCATGGCC--------TTACTTAC
ACTCCACCAG---------...35bp del...----------GAATGAATG
TGAGGTGGTC----------CTTACTTAC
TTATGCTGTGGTAAGAGTACCGGTG----------AATG
AATACGACACCATTCTCATGGCCAC----------TTAC
TTATGCTGTGGTAAGAGTACC--------AGGAATGAATG
AATACGACACCATTCTCATGG--------TCCTTACTTAC
TTAGGGGCTG----------...251bp del...----------AAATTAAAAT
AATCCCCGAC----------TTTAATTTTA

Figure 16. Targeted correction of the *F8* intron 22 locus in Hela cells. **(a)** Mutations at RGEN 02 and RGEN 03 target sites were estimated in Hela cells by the T7E1 assay. Asterisks indicate the predicted position of DNA bands cleaved by T7E1. **(b)** PCR products corresponding to inversion genotypes. **(c, d)** DNA sequences of two breakpoint junctions in transfected cells. The RGEN target sequence is underlined.

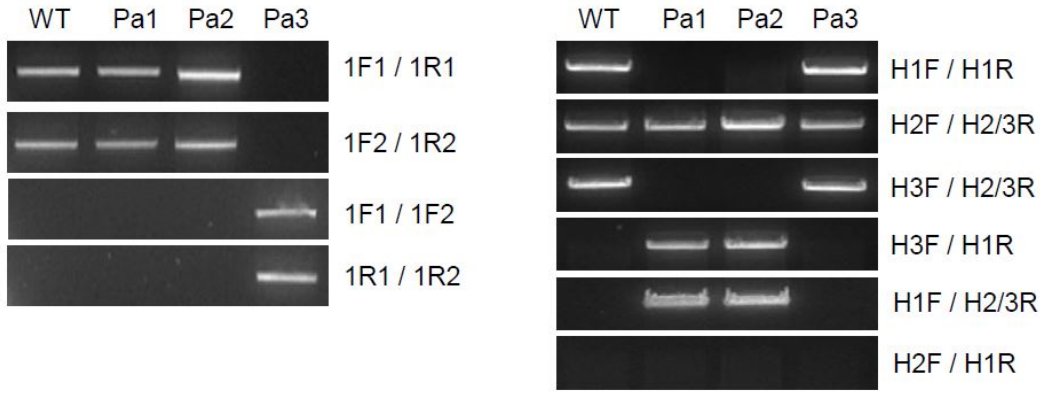


	Amount of genomic DNA(Copy number per half genome)					Frequency(%)		p value (Fit)
	1ng	333pg	100pg	33pg	10pg		Upper and	
	(300)	(100)	(30)	(10)	(3)	Estimate	Lower limits	
junction1	16/16	13/16	8/16	6/16	0/16	2.2	(1.4~3.3)	0.43
junction3	16/16	13/16	6/16	0/16		1.5	(1.0~2.3)	0.13

Figure 17. Frequencies of targeted intron22 inversion. The digital PCR was performed for estimating the frequency of targeted inversion as in **Figure 14.** Genomic DNA was isolated from the HeLa cells transfected with both of RGEN02 and 03 and was serially diluted. Estimated frequency of targeted inversions events created via RGENs was measured by digital PCR analysis. Upper and Lower limits indicate 95% confidence intervals.

3. Genotype of urine-derived HA patient cells

To create iPSCs from hemophilia A patients, the genotype screening is performed to search intron 1 and 22 inversions from 11 unrelated korean patient cells with severe hemophilia A. Only 1 patient has intron 1 inversion, and 3 patients have intron 22 inversion. These incidences found in here coincide approximately with previously reports (Lakich et al. 1993; Kim et al. 2012). After screening, 3 patient cells were selected (namely, Pa1, Pa2, and Pa3) for further investigation. As showed in Figure 18, Pa1 and Pa2 cells showed wild-type in intron-1 locus of *F8*, but Pa3 cells showed inversion in this locus. In contrast, Pa1 and Pa2 patient cells presented a genotype of intron-22 inversion, but Pa3 cells showed wild-type. These results show that each patient cells harbor only one among the intron-1 or intron-22 inversion. In addition, PCR amplicon used with specific primer set, H2F and H1R, does not showed in Pa1 and Pa2 cells (Figure 18). This means that those cells have the genomic inversion between homolog 1 and 3 among the three homologies not between homolog 1 and 2 (Figure 11).



(by Park CY. in Yonsei Univ.)

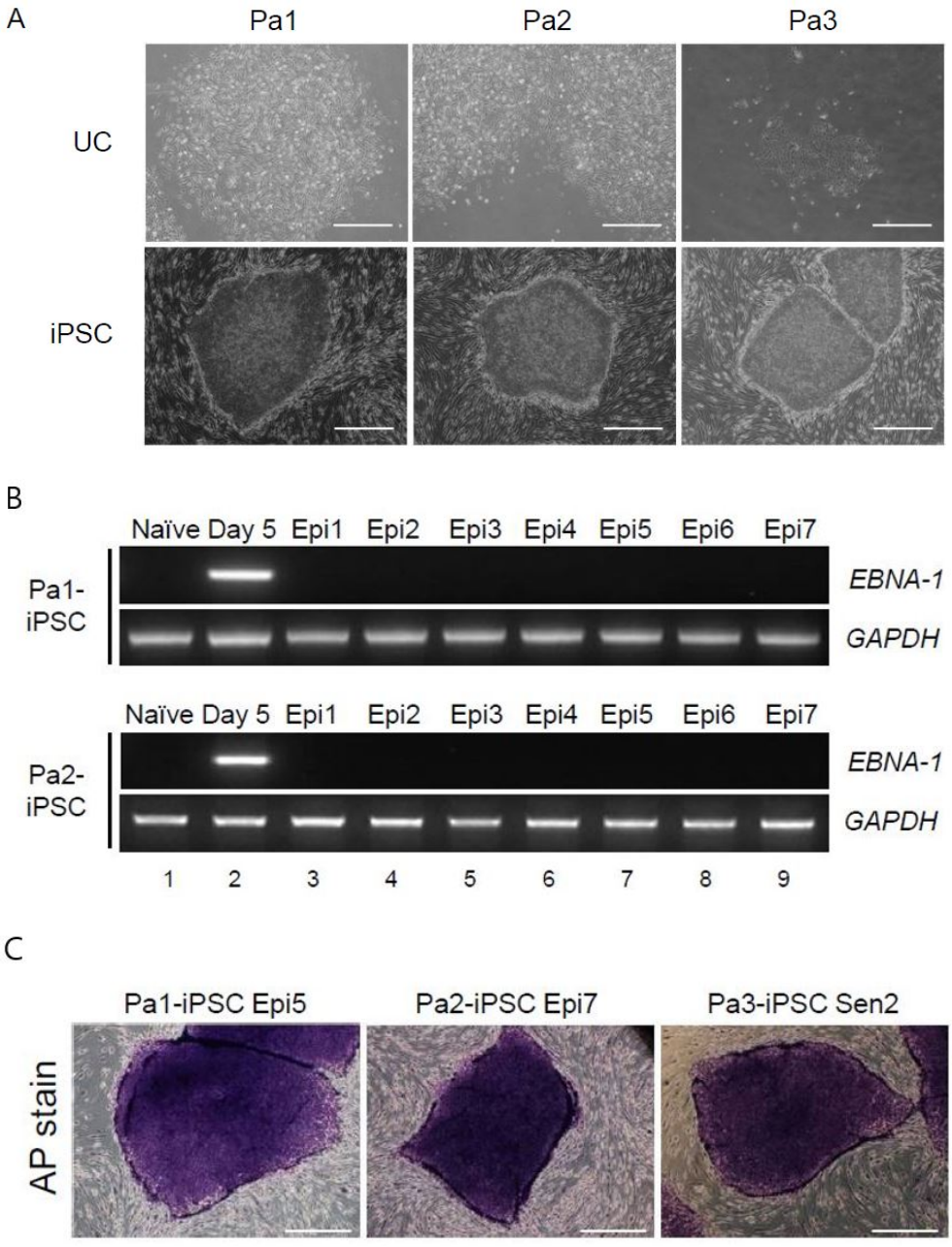
Figure 18. Genotype of the *F8* gene in patient cells. Genomic DNA was isolated from human dermal fibroblasts (WT) and hemophilia A patient-derived urine cells (Pa1, Pa2, and Pa3). PCR analysis was subjected to identify of the intron 1 (left) and 22 (right) inversions using the specific primers (listed in Table 4; Figure 11).

4. Generation of iPS cells from urine-derived patient cells

Next, each iPSCs from 3 urine-derived patient cells were generated (Figure 19). Through both method electroporation of episomal vector or infection of Sendai virus containing the reprogramming factors, it was able to obtain embryonic stem cell (ESC)-like colonies derived from urine cells (Figure 19A). Among them, total of 7 colonies (termed Epi1 to Epi7 in case of Pa1- and Pa2-iPSCs, respectively) derived by episomal vector from Pa1 and Pa2 urinary cells and total of 5 colonies (termed Sen1 to Sen5) derived by Sendai virus from Pa3 urinary cells were propagated. In case used episomal vector, the EBNA-1 sequence was amplified through PCR analysis to check the absence of residual vector in these lines after 8 passages. All lines selected were did not contained the EBNA-1 sequence (Figure 19B) and were showed high alkaline phosphatase activities (Figure 19C). After initial characterizations, one line from each patient iPSCs (namely, Epi5, Epi7, and Sen2 from Pa1-, Pa2-, and Pa3-iPSCs, respectively) is selected for further study. Each iPSC lines had a normal karyotype (Figure 19D). Immunostaining analysis revealed that the typical ESC marker proteins were well expressed in these lines (Figure 19E).

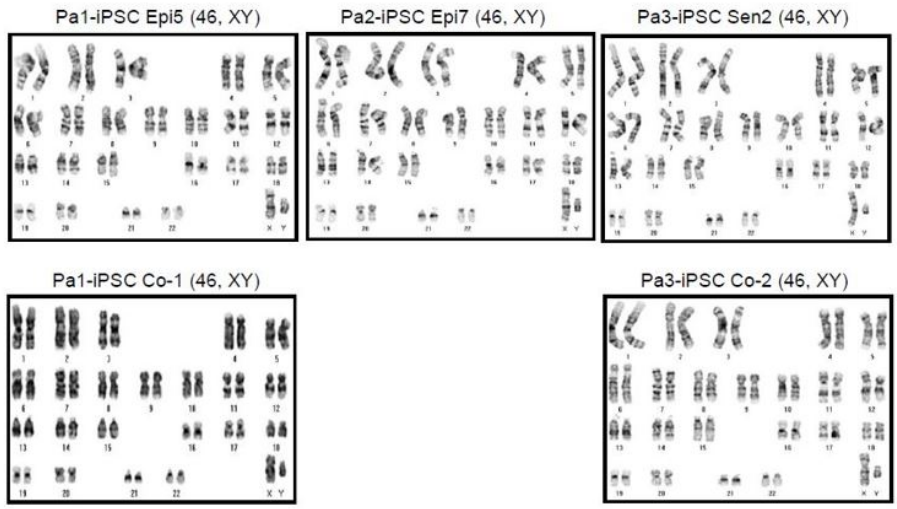
Next, the mRNA level of pluripotent marker genes are confirmed such as Oct4, Sox2, Lin28, and Nanog in selected lines. As shown in Figure 19F, those expressed at similar levels in iPSC lines with the human ES cell line, H9. In addition, in vitro differention potential of these lines was

investigated. It was found that marker proteins for three germ layers were detected in the differentiated cells by immunostaining analysis (Figure 19G). These data show that all selected lines derived from urinary cells of patient have pluripotency.

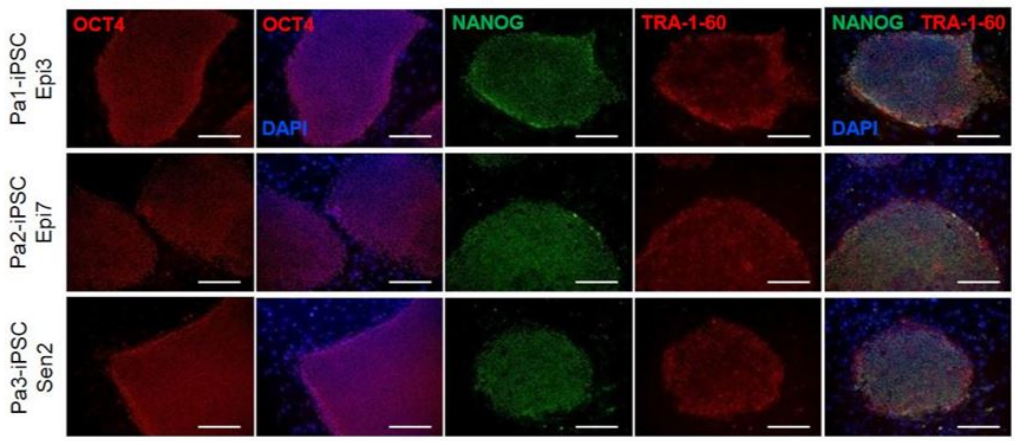

A
Pa1
Pa2
Pa3
UC
iPSC
B
Naïve Day 5 Epi1 Epi2 Epi3 Epi4 Epi5 Epi6 Epi7
Pa1-iPSC
EBNA-1
GAPDH
Naïve Day 5 Epi1 Epi2 Epi3 Epi4 Epi5 Epi6 Epi7
Pa2-iPSC
EBNA-1
GAPDH
1 2 3 4 5 6 7 8 9
C
Pa1-iPSC Epi5
Pa2-iPSC Epi7
Pa3-iPSC Sen2
AP stain

(continued)

D

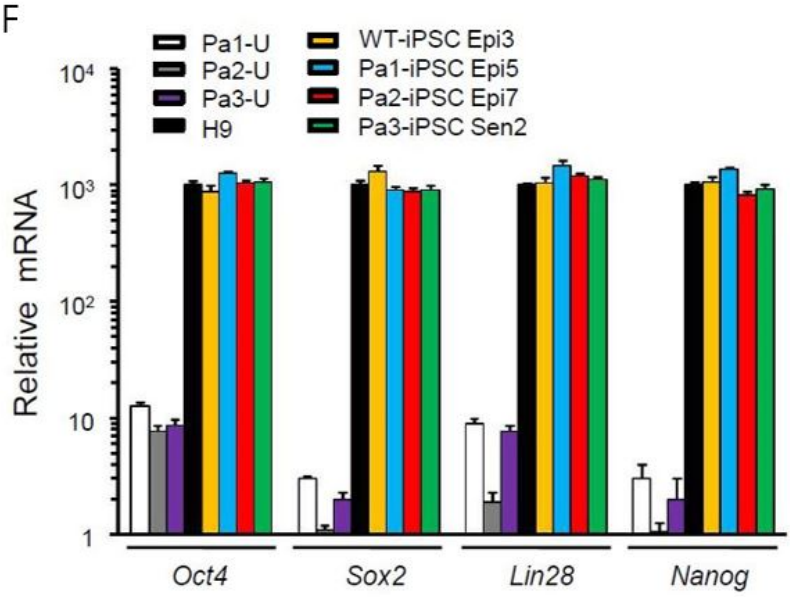


E

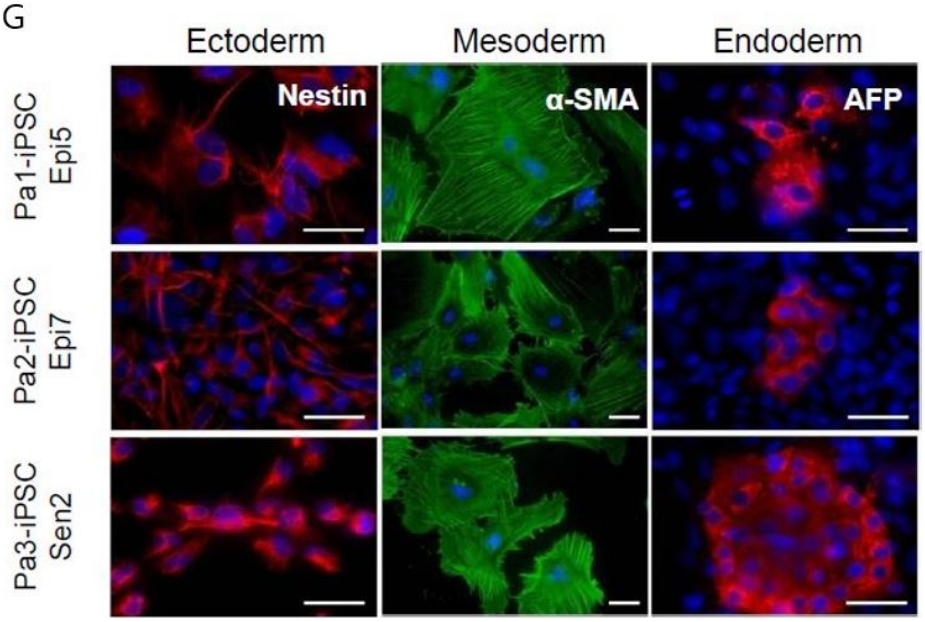


(continued)

F



G



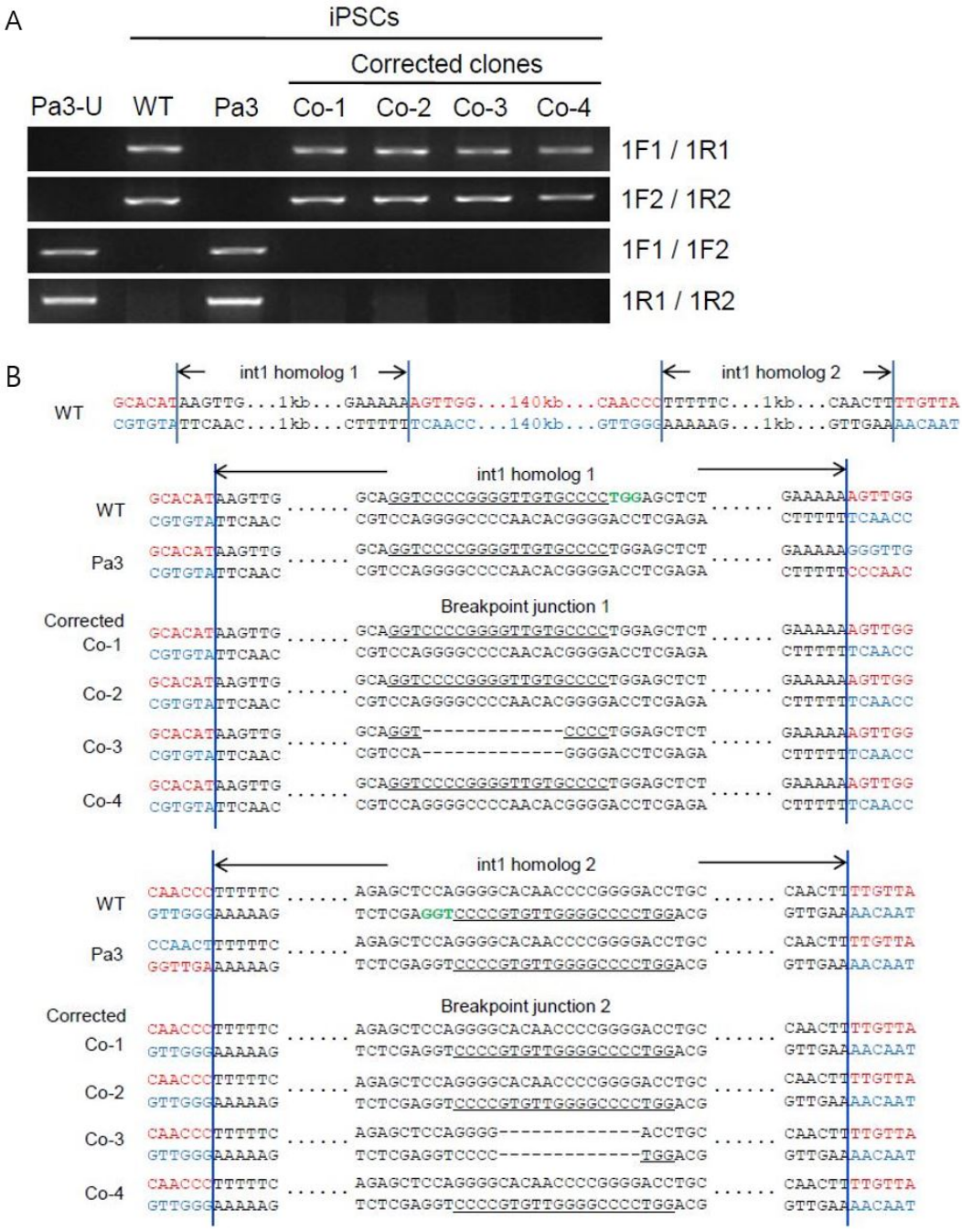
(by Park CY. in Yonsei Univ.)

Figure 19. Generation of HA patient-specific iPSCs from urine cells. **(a)** Morphology of the expanded urine cells (upper) and iPSC clones (lower). Scale bar, 500um. **(b)** Detection of an episomal vector sequence (EBNA-1) that remained in established iPSC lines (Epi1 to Epi7 of Pa1-iPSCs and Pa2-iPSCs, respectively). The GAPDH gene was used as a internal control for isolated total DNA. Total DNA isolated from the each urine cells before (naive) and after electroporation (Day 5) was used as negative and positive controls for episomal vector DNA. **(c)** Alkaline phosphatase staining of the indicated iPSCs lines. Scale bar, 500 um. **(d)** Karyotype analyses were performed on chromosomes from the indicated iPSC lines. **(e)** The expression of OCT4, NANOG and TRA-1-80, which are human ESC specific markers, was detected by immunocytochemistry. DAPI signals indicate the total cell presence in the image. Scale bar, 100um. **(f)** Quantitative real-time PCR (qPCR) was performed for detection of endogenous OCT4, SOX2, LIN28, and NANOG mRNAs in the indicated cell lines. The expression level of those genes was normalized to GAPDH expression. **(g)** Detection of marker proteins representing ectoderm (Nestin), mesoderm [α-smooth muscle actin (α-SMA)], and endoderm [α-fetoprotein (AFP)] in each lines by immunostaining with specific antibodies. DAPI signals indicate the total cell presence in the image. Scale bar, 50 um.

5. Targeted corrections of the *F8* gene from patient iPSCs with intron 1 inversions

To investigate availability of targeted inversion via RGENs as gene therapeutic way, RGEN01 was electroporated to repair of 140-kb inverted segment into the Pa3-iPSCs, and pulsed cells were further cultured to form colony. By PCR analysis, the specific bands for two homologies were detected using genomic DNA samples isolated from each colony. 8 colonies among total 120 colonies (6.7% frequency) have positive PCR bands for the two homologies by PCR screening. Four colonies were then further cultured to derive single cell clones. These corrected clones likewise wild-type cells produced homology specific PCR bands, while patient-derived urine cells and iPSCs did not produce these bands (Figure 20A).

Next, the sequence of these PCR products were analyzed to confirm the genotype. Indel mutations were not found in three clones, but 13-bp deletions at RGEN01 target sites in two homologies were found in the other clone (termed Co-3) (Figure 20B).

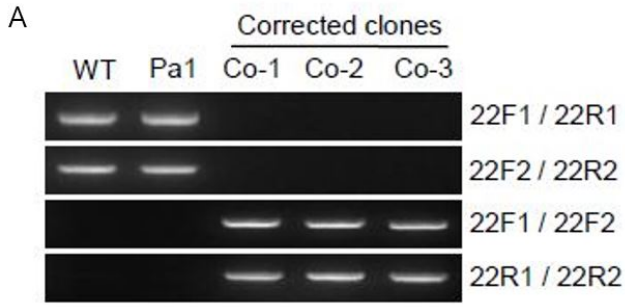


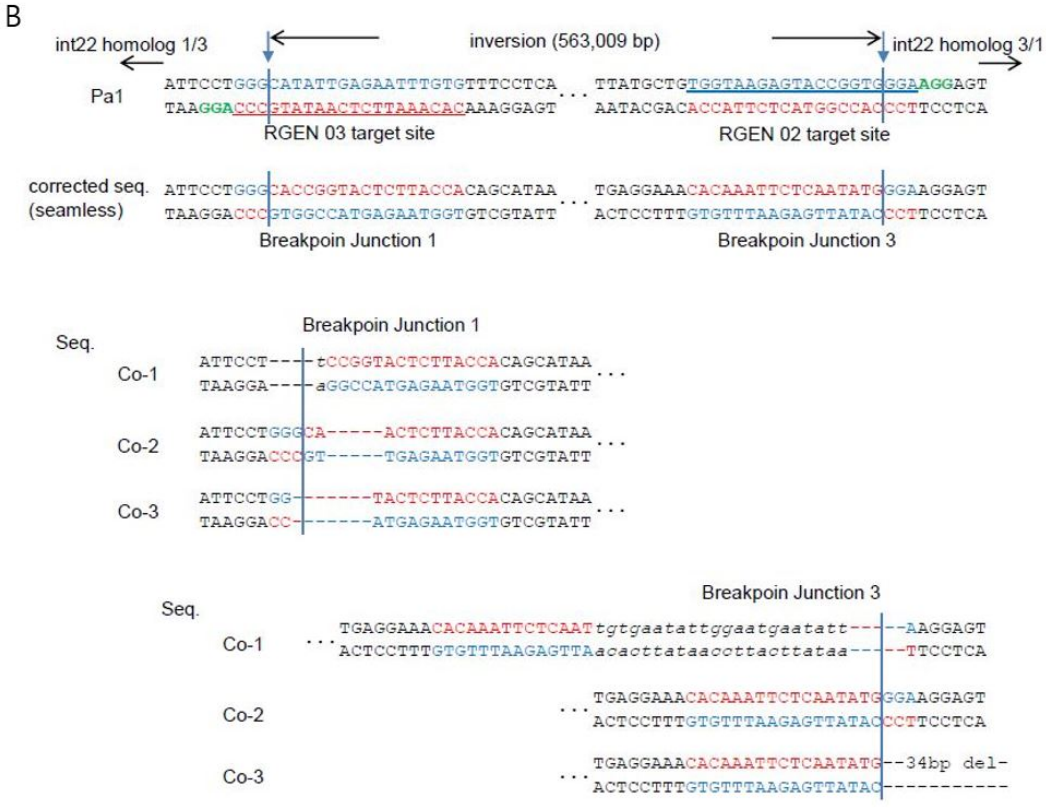
(by Park CY. in Yonsei Univ.)

Figure 20. Targeted correction of *F8* intron 1 locus from intron 1 inverted patient iPSCs. **(a)** PCR analysis of genomic DNA from four corrected clones (Co-1 to Co-4). Genomic DNAs were isolated from the urine-derived cells of intron 1 inversion patient (Pa3-U) or wild-type iPSCs (WT) served as positive controls for the inversion or normal genotypes, respectively. **(b)** DNA sequences of two homolog (here named int1 homolog 1 and 2) and breakpoint junctions at the indicated clones. The RGEN target sequence is underlined. Dashes indicate deleted bases.

6. Targeted corrections of the *F8* gene from patient iPSCs with intron 22 inversions

To introduce targeted re-inversion of intron22 in patient iPSCs, RGEN02 and RGEN03 were delivered in Pa1-iPS cells by electroporation. Five colonies of 135 (3.7% frequency) showed corrected PCR products for the two inversion breakpoint junctions. Three corrected clones were then derived after further culturing. By genotype PCR, these clones have corrected segment (namely, 600-kbp inverted between nearby int22h1 and int22h3), while patient-derived iPSCs involved intron 22 inversion do not have (Figure 21A). Small deletions or insertions were found at two RGEN target sites might be resulted by error-prone DSB repair via NHEJ (Figure 21B). These results show that 600-kbp inverted segment found in severe hemophilia A could be corrected using a couple of RGENs.





(by Park CY. in Yonsei Univ.)

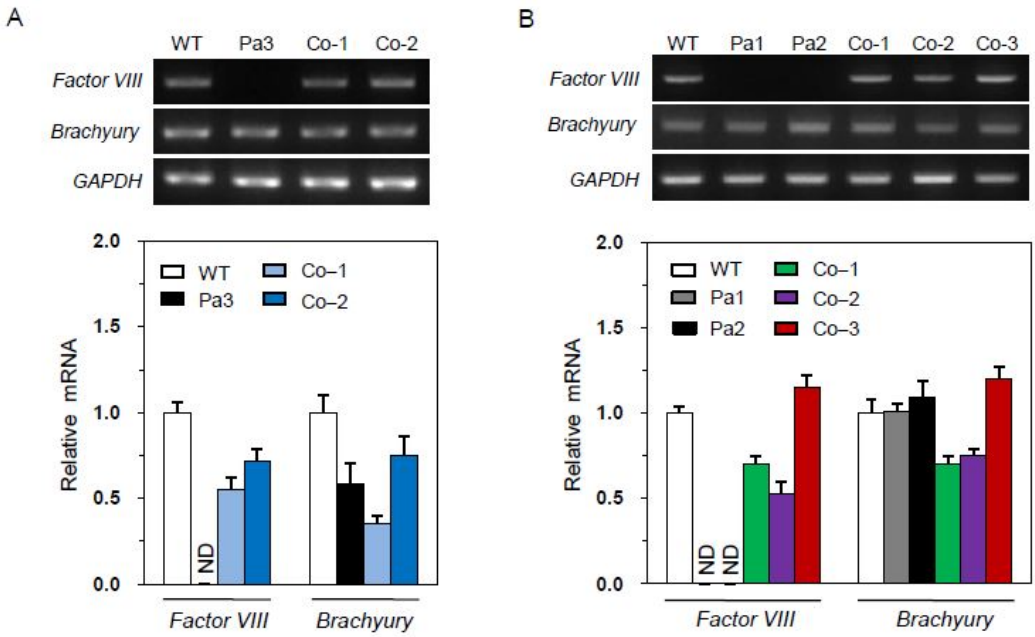
Figure 21. Targeted correction of *F8* intron 22 locus from intron 22 inverted patient iPSCs. **(a)** PCR analysis of genomic DNA from three corrected clones (Co-1 to Co-3). Genomic DNAs isolated from the wild-type iPSCs (WT) or intron 22 inverted patient iPSCs (Pa1) served as normal genotypes or positive controls for the inversion, respectively. **(b)** DNA sequences of breakpoint junctions in the indicated clones. Each RGEN target sequence is underlined, respectively. Dashes indicate deleted bases. A lowercase letter indicates inserted bases. Two blue arrows indicate cleavage sites.

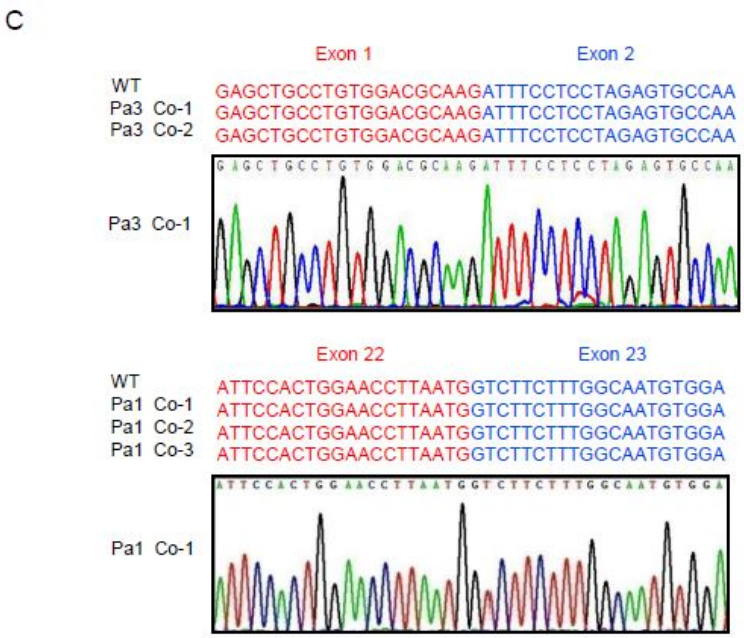
7. mRNA expression of *F8* in cells differentiated from corrected clones

It has been reported that endothelial cells which is derived from mesoderm were important source in the *F8* gene expression (Shahani et al. 2010). Therefore, the patient or corrected iPSCs were differentiated into mesoderm and investigated the expression of *F8* mRNA through an RT-PCR and qPCR analysis. First, it was determined whether amplicons of *F8* exon 1 to 2 could be detect in cells differentiated from Pa3-iPSCs with intron 1 inversion. As expected, no amplicons of *F8* exon 1 to 2 were detected in these cells, although the cells differentiated from Pa3-iPSCs could differentiate successfully into mesoderm as shown by the expression of Brachyury, a mesoderm cell marker gene (Figure 22). However, those amplicons were accurately detected in cells differentiated from the wild-type and two corrected lines (termed Co-1 and Co-2) repaired from Pa3-iPSCs.

Next, it was confirmed whether amplicons of *F8* exon 22 to 23 could be detect in cells differentiated from Pa1- and Pa2-iPSCs with intron 22 inversion. Amplicons of *F8* exon 22 to 23 were undetected in these cells, although the cells differentiated from those iPSCs could differentiate successfully into mesoderm (Figure 22). On the other hand, those amplicons were successfully detected in cells differentiated from the wild-type and three corrected lines (termed Co-1 to Co-3) repaired

from Pa1-iPSCs. By sequencing analysis, it was found that the splicing events between exon 1 and 2, or exon 22 and 23 from Pa3 and Pa1 corrected lines, respectively, could occur precisely without any unwanted genetic mutations (Figure 22C). These results prove that the integrity of the *F8* gene could correct accurately from patient iPSCs with intron 1 or 22 inversion, which supports expression of the *F8* gene in mesoderm cells.



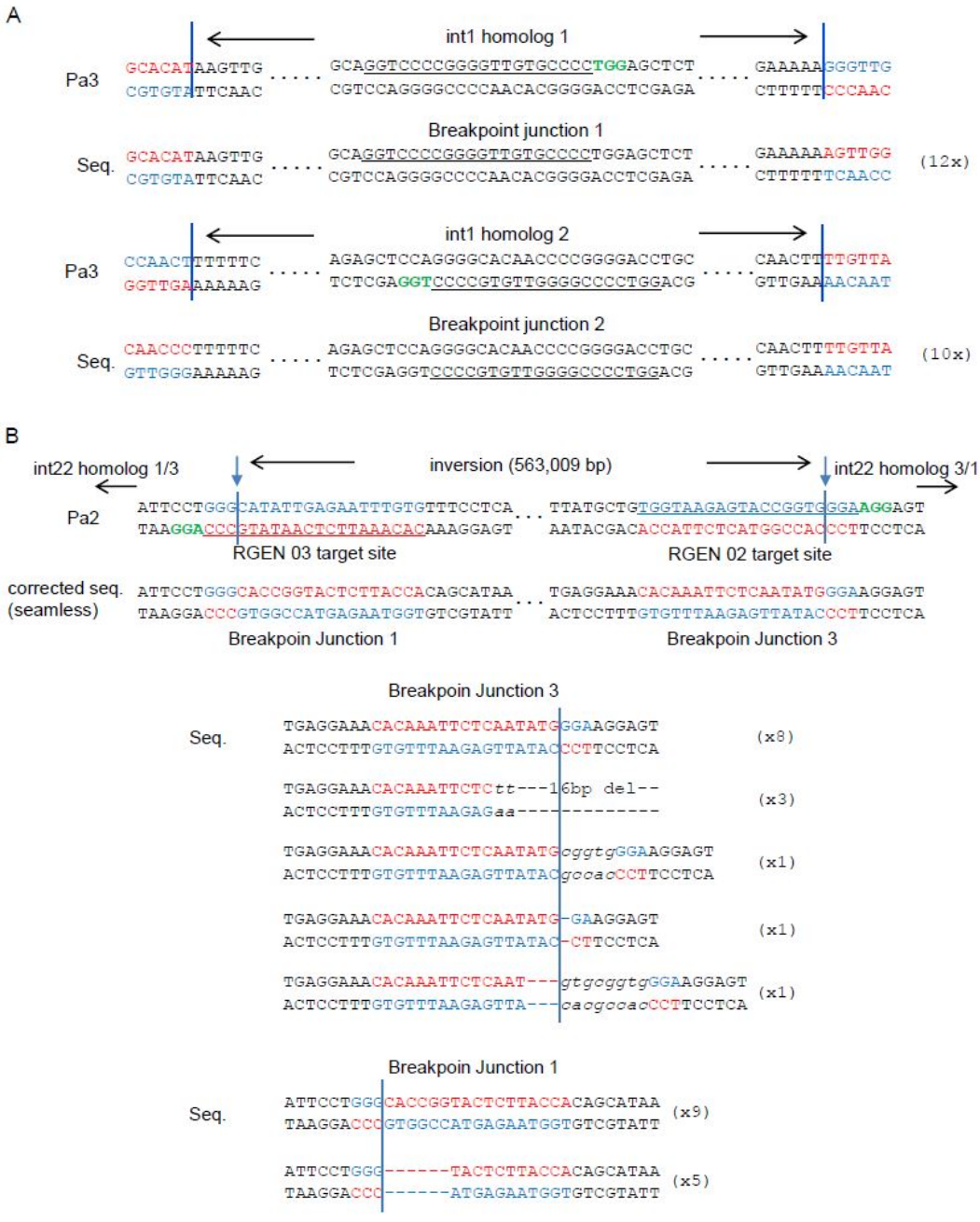


(by Park CY. in Yonsei Univ.)

Figure 22. Expression of the *F8* gene in corrected clones. **(a)** *F8* expression in cells differentiated from intron−1 corrected clones. RT−PCR (upper) and qPCR (lower) was used to detect expression of F8 and mesoderm marker gene (Brachyury) in cells derived from wild−type iPSCs (WT), intron−1 inverted patient iPSCs (Pa3), and corrected clones (Co−1 and Co−2). GAPDH expression was used as a loading control. **(b)** *F8* expression in cells derived from intron−22 corrected clones. RT−PCR (upper) and qPCR (lower) was used to detect expression of *F8* and mesoderm marker gene (Brachyury) in cells derived from wild−type iPSCs (WT), intron−22 inverted patients iPSCs (Pa1 and Pa2), and corrected clones (Co−1, Co−2 and Co−3) which are originated from Pa1−iPSCs. **(c)** The sequences between exon 1 to 2 or exon 22 to 23 in wild−type cells and corrected cells were verified, and the chromatograms were shown.

8. Targeted correction of the *F8* gene using RGEN ribonucleoproteins (RNPs) in patient-iPSCs

Recently, it has been reported that RGEN RNPs can introduce into the fibroblast and ES cells via direct delivery and induce efficiently indels at target sites (Kim et al. 2014). Because this method has advantage in safety aspect than plasmid encoding RGENs, this method was also used for correction of the *F8* gene from patient-iPSCs. The RGEN RNPs were transfected into two patient iPSCs containing intron 1 or intron 22 inversion (Pa3- and Pa2-iPSCs, resfectively), and then total genomic DNA samples were isolated at 4 day after transfection. The correction-specific PCR band was cloned and sequenced. As expected, precisely corrected sequence was found with no indels in Pa3-iPSCs with intron 1 inversion (Figure 23), suggesting the event of NAHR in RGEN-mediated genomic correction. Moreover, repaired sequences of the 600-kb inverted segment in *F8* locus were showed in Pa2-iPSCs with intron 22 inversion via a direct delivery of RGEN RNPs (Figure 23). The joining seamlessly and indels at the each breakpoint junction were revealed by sequence analysis, suggesting the involvement of NHEJ in this case.



(by Park CY. in Yonsei Univ.)

Figure 23. Targeted correction of the *F8* gene using RGEN ribonucleoproteins (RNPs) in patient-specific iPSCs. DNA sequences of each breakpoint junctions for the inversion events in the intron 1 **(a)** or intron 22 **(b)** regions. A direct delivery of RGEN RNPs was performed by electroporation into the intron 1 (Pa3-iPSCs) or intron 22 (Pa2-iPSCs) inverted cells. Total DNA isolated from the each iPSCs and sequenced. Each RGEN target sequence is underlined, respectively. Dashes indicate deleted bases. A lowercase letter indicates inserted bases. Two blue arrows indicate cleavage sites. In cases in which a sequence was detected more than once, the number of occurrences is shown in parentheses.

9. Analysis of RGEN off-target effect

In recent years, many groups reported that RGENs had off-target cleavage effect in human cells (Fu et al. 2013; Hsu et al. 2013; Pattanayak et al. 2013). Off-target cleavage can be induced at sites that are highly homologous to on-target sites. As mutations in off-target sites are very serious problem especially in gene therapy, I tested whether there are indel mutations at potential off-target sites induced by NHEJ with off-target cleavage of RGENs. Potential off-target sites of RGENs used in this study were searched *in silico* (Bae et al. 2014; Table 5). T7E1 assay of those sites showed that there is no mutations at potential off-target sites of three RGENs in Hela cells treated with each RGENs (Figure 24).

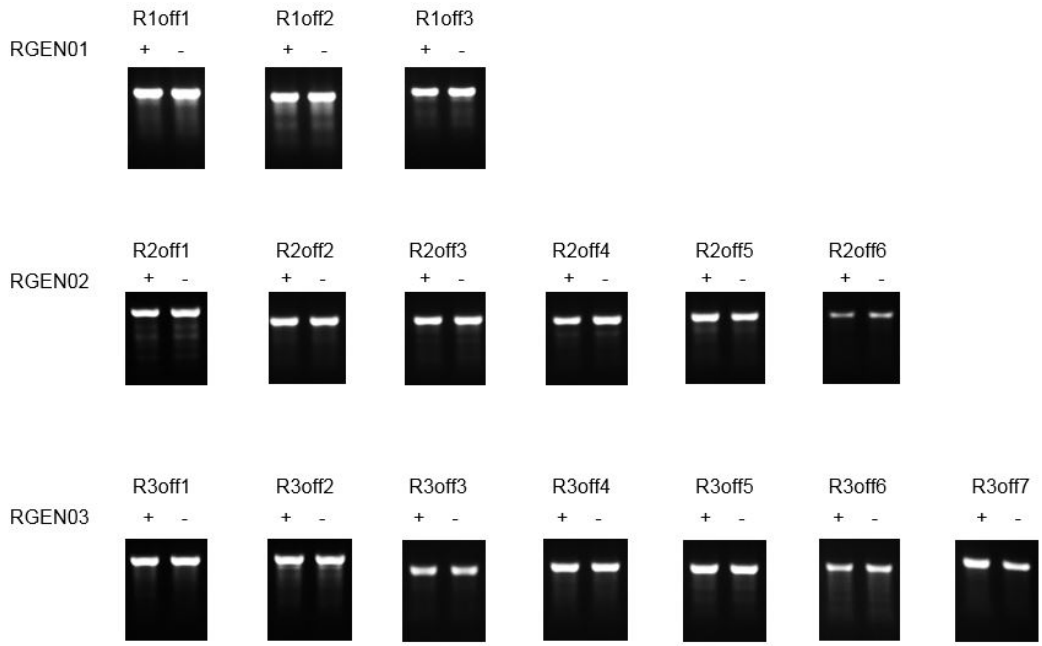


Figure 24. Analysis of RGEN off-target effect. Potential off-target sites of RGENs used in this study were searched *in silico* (Bae et al. 2014). RGEN-encoding plasmids were transfected in HeLa cells. Several potential off-target sites that similar to the RGEN target site were selected and amplified by PCR. The T7E1 analysis was performed to confirm the off-target cleavage activities at these sites.

Table 5. Potential off-target sites of RGENs in this study. Identical nucleotides are in uppercase. PAM, Protospacer Adjacent Motif.

Locus	Seqeunces	PAM
RGEN01	GGTCCCCGGGGTTGTGCCCC	TGG
R1off1	tGgCCCCtGGGTTGTGCCCC	AGG
R1off2	GGTCCCtGGGGTgtTGCCCC	TGG
R1off3	GGgCCCCGGGGTccTGCCCC	AGG
RGEN02	TGGTAAGAGTACCGGTGGGA	AGG
R2off1	TaGTAAtAGTAaCGGTGGGA	GGG
R2off2	TGGTgAGAGTAggGGTGGGA	GGG
R2off3	TGGaAAGAGaACCGGTGtGA	AGG
R2off4	TGGggAGAGTACCGGcGGGA	AGG
R2off5	TGGTcAGAGTAaCaGTGGGA	TGG
R2off6	TGGTAAaAGTAaCtGTGGGA	TGG
RGEN03	CACAAATTCTCAATATGCCC	AGG
R3off1	CcCAAATcCTCAATATGCCC	CGG
R3off2	CACAAATTCTCAATtTGCaC	AGG
R3off3	CACAAAgcCTCAAcATGCCC	TGG
R3off4	gAaAAATTCTaAATATGCCC	TGG
R3off5	tACAAATTtTCcATATGCCC	TGG
R3off6	aACAtATTCTCAAaATGCCC	AGG
R3off7	CACAAAaTCaaAATATGCCC	TGG

Discussion

Hemophilia A is one of the most common genetic bleeding disorders caused by various mutations in the *F8* gene. In this study, I designed CRISPR/Cas system as engineered nucleases targeting inverted sites in the *F8* gene. These RGENs can induce double strand break at target sites very efficiently (up to 44%) in Hela cells tested by T7E1 assay. Targeted inversion via these RGENs is also very efficient (up to 6.7%) to generate single clone with ease. Importantly, corrected clones from severe hemophilia A patient iPSCs can regenerate F8 mRNA like as wild type cells.

Induced DSB via programmable endonucleases at precise target loci can be repaired by non-homologous end-joining (NHEJ) or homologous recombination (HR) resulting mutations and chromosomal rearrangements (Bibikova et al. 2003; Kim et al. 2009; Kim et al. 2011b; Miller et al. 2011; Brunet et al. 2009; Lee et al. 2010; Lee et al. 2012). Thus, engineered nucleases are emerging as powerful tools for gene therapy. Although ZFNs and TALENs were well studied artificial nucleases (Kim et al. 2009; Kim et al. 2011b; Reyon et al. 2012; Kim et al. 2013), RGEN is more easy method to design and customizing. To create new RGEN, only expression vector of sgRNA should be replaced by cloning with 20-bp target sequence without changing the Cas9 protein component. Furthermore, for making RNP complex to deliver into cells

which is a novel method reducing off-target cleavage (Kim et al. 2014), Cas9 protein should not be purified every time because target recognition function fully relies on the guide RNA. Actually, RGEN system is broadly used in cells (Cho et al. 2013a; Cong et al. 2013; Jinek et al. 2013) and many species (Cho et al. 2013b; Hwang et al. 2013; Sung et al. 2014; Dickinson et al. 2013; Friedland et al. 2013; Gratz et al. 2013; Li et al. 2013b; Nekrasov et al. 2013; Shan et al. 2013).

As previously reported, targeted cleavage at two endogenous loci in the genome of human cells can result in deletions, inversions and duplications (Lee et al. 2011). In previous report and this study, inversion of intron1 in the *F8* gene is corrected by reversion of segment between intron1 and corresponding homologous sequences located far upstream of the *F8* gene (Park et al. 2014). Frequent detection of seamless junction in cleavage sites imply this junction might be repaired by NAHR more than error prone NHEJ.

Inversion of intron22 in the *F8* gene is more complicated. Because there are two more homologous sequences corresponding intron22 in the upstream region of the *F8* gene, targeted cleavage inside homologous sequences result in DSB at three loci total and, as a result, two genomic fragments are created. As repair of two genomic fragments lead to many combination of deletions, inversions and duplications, the chance to get perfectly corrected clones of intron22 inversion is probably very low. Actually, I designed RGEN targeting inside of homology of intron22, but

corrected clones had not been created because of low efficiency. Instead, by targeting the locus a bit aside of homology, I could induce cleavage at two sites in the genome, and efficiently create corrected clone of intron22 in patient iPSC whose expression of *F8* mRNA is normal like wild-type cells although DNA sequence nearby junction is different from that of wild type cells.

Targeting with RGEN in non-homologous sequences also has a benefit on detection of inversion. Homologous sequence of intron22 reaches a length of 10kb and additionally 50kb in case of between *int22h2* and *int22h3,* indeed, standard PCR method cannot be used for genotyping. Although inverse shifting-polymerase chain reaction (IS-PCR) approach can detect inversion of intron22 (Rossetti et al. 2008), this method is not suitable for early screening of iPSC clones because it needs lots of genomic DNA (~2ug). In this study, two RGENs target unique sites individually and, consequently, corrected clones were selected by standard PCR with ease.

In summary, RGENs can be used to induce targeted inversions efficiently in Hela cells and human iPSCs that represent severe hemophilia A. Inverted 140kbp genomic DNA segment of intron1 inversion patient cells, and about 600kbp that of intron22 inversion patient cells were successfully reverted via induced DBS at precise target loci. Furthermore, mRNA expression of the *F8* gene in corrected clones is recovered like as normal cells. Although *in vivo* study of these corrected clones is not yet

demonstrated, this study showed that artificial inversion approach via engineered nucleases can be used for autologous stem cell therapy to treat severe hemophilia A patients. RGEN-induced genomic rearrangements will not only suggest a superb gene therapeutic approach for many genetic diseases but provide an experimental platform to study the molecular mechanism and biological functions of genomic structures.

References

Adenot PG, Mercier Y, Renard JP, Thompson EM. 1997. Differential H4 acetylation of paternal and maternal chromatin precedes DNA replication and differential transcriptional activity in pronuclei of 1-cell mouse embryos. *Development*. **124**(22):4615-4625.

Bae S, Park J, Kim JS. 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics.* **30**(10):1473-1475.

Bagnall RD, Waseem N, Green PM, Giannelli F. 2002. Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A. *Blood.* **99**(1):168-174.

Bagnall RD, Giannelli F, Green PM. 2006. Int22h-related inversions causing hemophilia A: a novel insight into their origin and a new more discriminant PCR test for their detection. *J Thromb Haemost.* **4**(3):591-598.

Beumer KJ, Trautman JK, Bozas A, Liu JL, Rutter J, Gall JG, Carroll D. 2008. Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. *Proc Natl Acad Sci U S A.* **105**(50):19821-19826.

Bibikova M, Beumer K, Trautman JK, Carroll D. 2003. Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**(5620):764.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*. **326**(5959):1509-1512.

Boch J, Bonas U. 2010. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol*. **48**:419-436

Bogdanove AJ, Schornack S, Lahaye T. 2010. TAL effectors: finding plant genes for disease and defense. *Curr Opin Plant Biol*. **13**(4):394-401.

Brunet E, Simsek D, Tomishima M, DeKelver R, Choi VM, Gregory P, Urnov F, Weinstock DM, Jasin M. 2009. Chromosomal translocations induced at specified loci in human stem cells. *Proc Natl Acad Sci U S A.* **106**(26):10620-10625.

Cappelli B, Aiuti A. 2010. Gene therapy for adenosine deaminase deficiency. *Immunol Allergy Clin North Am.* **30**(2):249-260.

Carbery ID, Ji D, Harrington A, Brown V, Weinstein EJ, Liaw L, Cui X. 2010. Targeted genome modification in mice using zinc-finger nucleases. *Genetics* **186**(2):451-459.

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39**(12):e82.

Cho SW, Kim S, Kim JM, Kim JS. 2013a. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol.* **31**(3):230-232.

Cho SW, Lee J, Carroll D, Kim JS, Lee J. 2013b. Heritable gene knockout in Caenorhabditis elegans by direct injection of Cas9-sgRNA ribonucleoproteins. *Genetics.* **195**(3):1177-1180.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science.* **339**(6121):819-823.

Cornu TI, Thibodeau-Beganny S, Guhl E, Alwin S, Eichtinger M, Joung JK, Cathomen T. 2008. DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol Ther.* **16**(2):352-358.

Desbordes SC, Placantonakis DG, Ciro A, Socci ND, Lee G, Djaballah H, Studer L. 2008. High-throughput screening assay for the identification of compounds regulating self-renewal and differentiation in human embryonic stem cells. *Cell Stem Cell.*

2(6):602-612.

Dickinson DJ, Ward JD, Reiner DJ, Goldstein B. 2013. Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. *Nat Methods.* 10(10):1028-1034.

Doyle EL, Booher NJ, Standage DS, Voytas DF, Brendel VP, Vandyk JK, Bogdanove AJ. 2012. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.* 40(Web Server issue):W117-122.

Doyon Y, McCammon JM, Miller JC, Faraji F, Ngo C, Katibah GE, Amora R, Hocking TD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Amacher SL. 2008. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol.* 26(6):702-708.

Doyon Y, Vo TD, Mendel MC, Greenberg SG, Wang J, Xia DF, Miller JC, Urnov FD, Gregory PD, Holmes MC. 2011. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods.* 8(1):74-79.

Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411(6836), 494-498.

Feuk L, Carson AR, Scherer SW. 2006. Structural variation in the human

genome. *Nat Rev Genet.* **7**(2):85-97.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**(6669):806-811.

Friedland AE, Tzur YB, Esvelt KM, Colaiácovo MP, Church GM, Calarco JA. 2013. Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nat Methods.* **10**(8):741-743.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol.* **31**(9):822-826.

Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM. 2013. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics.* **194**(4):1029-1035.

Graw J, Brackmann HH, Oldenburg J, Schneppenheim R, Spannagl M, Schwaab R. 2005. Haemophilia A: from mutation analysis to new therapies. *Nat Rev Genet.* **6**(6):488-501.

Guo J, Gaj T, Barbas CF 3rd. 2010. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J Mol Biol.* **400**(1):96-107

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol.* **31**(3):227-229.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol.* **31**(9):827-832.

Hu Y, Smyth GK. 2009. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J Immunol Methods.* **347**(1-2):70-78.

Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, Li B, Cavet G, Linsley PS. 2003. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol.* **21**(6):635-637. Epub 2003 May 18.

Jang J, Kang HC, Kim HS, Kim JY, Huh YJ, Kim DS, Yoo JE, Lee JA, Lim B, Lee J, Yoon TM, Park IH, Hwang DY, Daley GQ, Kim DW. 2011. Induced pluripotent stem cell models from X-linked adrenoleukodystrophy patients. *Ann Neurol.* **70**(3):402-409.

Jang J, Yoo JE, Lee JA, Lee DR, Kim JY, Huh YJ, Kim DS, Park CY, Hwang DY, Kim HS, Kang HC, Kim DW. 2012. Disease-specific induced pluripotent stem cells: a platform for human disease modeling and

drug discovery. *Exp Mol Med.* **44**(3):202-213.

Jeggo PA. 1998. Identification of genes involved in repair of DNA double-strand breaks in mammalian cells. 1998. *Radiat Res.* **150**(5 Suppl):S80-91.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* **337**(6096):816-821.

Kasparek P, Krausova M, Haneckova R, Kriz V, Zbodakova O, Korinek V, Sedlacek R. 2014. Efficient gene targeting of the Rosa26 locus in mouse zygotes using TALE nucleases. *FEBS Lett.* **S0014-5793**(14)00689-9.

Kim DS, Lee JS, Leem JW, Huh YJ, Kim JY, Kim HS, Park IH, Daley GQ, Hwang DY, Kim DW. 2010. Robust enhancement of neural differentiation from human ES and iPS cells regardless of their innate difference in differentiation propensity. *Stem Cell Rev.* **6**(2):270-281.

Kim H, Um E, Cho SR, Jung C, Kim H, Kim JS. 2011a. Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat Methods.* **8**(11):941-943

Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS. 2009. Targeted genome editing in human cells with zinc finger nucleases constructed via modular

assembly. *Genome Res.* **19**(7):1279-1288.

Kim HJ, Chung HS, Kim SK, Yoo KY, Jung SY, Park IA, Lee KO, Kim SH, Kim HJ. 2012. Mutation spectrum and inhibitor risk in 100 Korean patients with severe haemophilia A. 2012. *Haemophilia*. **18**(6):1008-1013.

Kim S, Lee MJ, Kim H, Kang M, Kim JS. 2011b. Preassembled zinc-finger arrays for rapid construction of ZFNs. *Nat Methods.* **8**(1):7

Kim S, Kim D, Cho SW, Kim J, Kim JS. 2014. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**(6):1012-1019.

Kim Y, Kweon J, Kim A, Chon JK, Yoo JY, Kim HJ, Kim S, Lee C, Jeong E, Chung E, Kim D, Lee MS, Go EM, Song HJ, Kim H, Cho N, Bang D, Kim S, Kim JS. 2013. A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol.* **31**(3):251-258.

Kim YG, Cha J, Chandrasegaran S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A.* **93**(3):1156-1160.

Krueger U, Bergauer T, Kaufmann B, Wolter I, Pilk S, Heider-Fabian M, Kirch S, Artz-Oppitz C, Isselhorst M, Konrad J. 2007. Insights into effective RNAi gained from large-scale siRNA validation screening. *Oligonucleotides* **17**(2):237-250.

Lakich D, Kazazian HH Jr, Antonarakis SE, Gitschier J. 1993. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat Genet.* **5**(3):236-241.

Lam KN, van Bakel H, Cote AG, van der Ven A, Hughes TR. 2011. Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. *Nucleic Acids Res.* **39**(11):4680-4690

Lee HJ, Kim E, Kim JS. 2010. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res.* **20**(1):81-89.

Lee HJ, Kweon J, Kim E, Kim S, Kim JS. 2012. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res.* **22**(3):539-548.

Li L, Krymskaya L, Wang J, Henley J, Rao A, Cao LF, Tran CA, Torres-Coronado M, Gardner A, Gonzalez N, Kim K, Liu PQ, Hofer U, Lopez E, Gregory PD, Liu Q, Holmes MC, Cannon PM, Zaia JA, DiGiusto DL. 2013a. Genomic editing of the HIV-1 coreceptor CCR5 in adult hematopoietic stem and progenitor cells using zinc finger nucleases. *Mol Ther.* **21**(6):1259-1269.

Li D, Qiu Z, Shao Y, Chen Y, Guan Y, Liu M, Li Y, Gao N, Wang L, Lu X, Zhao Y, Liu M. 2013b. Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat Biotechnol.* **31**(8):681-683.

Lloyd A, Plaisier CL, Carroll D, Drews GN. 2005. Targeted mutagenesis

using zinc-finger nucleases in Arabidopsis. *Proc Natl Acad Sci U S A* **102**(6):2232-2237.

Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, Unger-Wallace E, Sander JD, Müller-Lerch F, Fu F, Pearlberg J, Göbel C, Dassie JP, Pruett-Miller SM, Porteus MH, Sgroi DC, Iafrate AJ, Dobbs D, McCray PB Jr, Cathomen T, Voytas DF, Joung JK. 2008. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell*. **31**(2):294-301

Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA. 2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol*. **26**(6):695-701.

Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, Beausejour CM, Waite AJ, Wang NS, Kim KA, Gregory PD, Pabo CO, Rebar EJ. 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*. **25**(7):778-785.

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD, Rebar

EJ. 2011. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol.* **29**(2):143-148.

Moskalenko S, Henry DO, Rosse C, Mirey G, Camonis JH, White MA. 2002. The exocyst is a Ral effector complex. *Nat Cell Biol.* **4**(1):66-72.

Nathwani AC, Tuddenham EG, Rangarajan S, Rosales C, McIntosh J, Linch DC, Chowdary P, Riddell A, Pie AJ, Harrington C, O'Beirne J, Smith K, Pasi J, Glader B, Rustagi P, Ng CY, Kay MA, Zhou J, Spence Y, Morton CL, Allay J, Coleman J, Sleep S, Cunningham JM, Srivastava D, Basner-Tschakarjan E, Mingozzi F, High KA, Gray JT, Reiss UM, Nienhuis AW, Davidoff AM. 2011. Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N Engl J Med.* **365**(25):2357-2365.

Naylor J, Brinke A, Hassock S, Green PM, Giannelli F. 1993. Characteristic mRNA abnormality found in half the patients with severe haemophilia A is due to large DNA inversions. *Hum Mol Genet.* **2**(11):1773-1778.

Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S. 2013. Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nat Biotechnol.* **31**(8):691-693

Okita K, Matsumura Y, Sato Y, Okada A, Morizane A, Okamoto S, Hong H, Nakagawa M, Tanabe K, Tezuka K, Shibata T, Kunisada T,

Takahashi M, Takahashi J, Saji H, Yamanaka S. 2011. A more efficient method to generate integration-free human iPS cells. *Nat Methods.* **8**(5):409-412.

Ohta T, Essner R, Ryu JH, Palazzo RE, Uetake Y, Kuriyama R. 2002. Characterization of Cep135, a novel coiled-coil centrosomal protein involved in microtubule organization in mammalian cells. *J Cell Biol.* **156**(1):87-99.

Park CY, Kim J, Kweon J, Son JS, Lee JS, Yoo JE, Cho SR, Kim JH, Kim JS, Kim DW. 2014. Targeted inversion and reversion of the blood coagulation factor 8 gene in human iPS cells using TALENs. *Proc Natl Acad Sci U S A.* **111**(25):9253-9258.

Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol.* **31**(9):839-843.

Pruett-Miller SM, Connelly JP, Maeder ML, Joung JK, Porteus MH. 2008. Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. *Mol Ther.* **16**(4):707-717.

Radecke S, Radecke F, Cathomen T, Schwarz K. 2010. Zinc-finger nuclease-induced gene repair with oligodeoxynucleotides: wanted and unwanted target locus modifications. *Mol Ther.*

18(4):743-753.

Ramirez CL, Foley JE, Wright DA, Müller-Lerch F, Rahman SH, Cornu TI, Winfrey RJ, Sander JD, Fu F, Townsend JA, Cathomen T, Voytas DF, Joung JK. 2008. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods*. 5(5):374-375.

Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK. 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol*. 30(5):460-465.

Rossetti LC, Radic CP, Larripa IB, De Brasi CD. 2008. Developing a new generation of tests for genotyping hemophilia-causative rearrangements involving int22h and int1h hotspots in the factor VIII gene. *J Thromb Haemost.* 6(5):830-836.

Sander JD, Cade L, Khayter C, Reyon D, Peterson RT, Joung JK, Yeh JR. 2011. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol.* 29(8):697-698.

Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A, Collingwood TN. 2008. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A*. 105(15):5809-5814.

Shahani T, Lavend'homme R, Luttun A, Saint-Remy JM, Peerlinck K,

Jacquemin M. 2010. Activation of human endothelial cells from specific vascular beds induces the release of a FVIII storage pool. *Blood.* **115**(23):4902-4909.

Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, Zhang K, Liu J, Xi JJ, Qiu JL, Gao C. 2013. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol.* **31**(8):686-688.

Shin J, Chen J, Solnica-Krezel L. 2014. Efficient homologous recombination-mediated genome engineering in zebrafish using TALE nucleases. *Development.* **141**(19):3807-3818.

Stankiewicz P, Lupski JR. 2010. Structural variation in the human genome and its role in disease. *Annu Rev Med.* **61**:437-455.

Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, Greenman CD, Jia M, Latimer C, Teague JW, Lau KW, Burton J, Quail MA, Swerdlow H, Churcher C, Natrajan R, Sieuwerts AM, Martens JW, Silver DP, Langerød A, Russnes HE, Foekens JA, Reis-Filho JS, van 't Veer L, Richardson AL, Børresen-Dale AL, Campbell PJ, Futreal PA, Stratton MR. 2009. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature.* **462**(7276):1005-1010.

Sugii S, Kida Y, Kawamura T, Suzuki J, Vassena R, Yin YQ, Lutz MK,

Berggren WT, Izpisúa Belmonte JC, Evans RM. 2010. Human and mouse adipose-derived cells support feeder-independent induction of pluripotent stem cells. *Proc Natl Acad Sci U S A.* **107**(8):3558-3563.

Sung YH, Baek IJ, Kim DH, Jeon J, Lee J, Lee K, Jeong D, Kim JS, Lee HW. 2013. Knockout mice created by TALEN-mediated gene targeting. *Nat Biotechnol.* **31**(1):23-24.

Sung YH, Kim JM, Kim HT, Lee J, Jeon J, Jin Y, Choi JH, Ban YH, Ha SJ, Kim CH, Lee HW, Kim JS. 2014. Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. *Genome Res.* **24**(1):125-131.

Szczepek M, Brondani V, Büchel J, Serrano L, Segal DJ, Cathomen T. 2007. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol.* **25**(7):786-793.

Katsuyama T, Akmammedov A, Seimiya M, Hess SC, Sievers C, Paro R. 2013. An efficient strategy for TALEN-mediated genome engineering in Drosophila. *Nucleic Acids Res.* **41**(17):e163.

Tadros W, Lipshitz HD. The maternal-to-zygotic transition: a play in two acts. *Development.* **136**(18):3033-3042.

Tesson L, Usal C, Ménoret S, Leung E, Niles BJ, Remy S, Santiago Y, Vincent AI, Meng X, Zhang L, Gregory PD, Anegon I, Cost GJ. 2011. Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol*. **29**(8):695-696.

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 2005 Jun 2;435(7042):646-651.

van Gent DC, Hoeijmakers JH, Kanaar R. 2001. Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet*. **2**(3):196-206

Wiedenheft B, Sternberg SH, Doudna JA. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature.* **482**(7385):331-338.

Wright DA, Thibodeau-Beganny S, Sander JD, Winfrey RJ, Hirsh AS, Eichtinger M, Fu F, Porteus MH, Dobbs D, Voytas DF, Joung JK. 2006. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc*. **1**(3):1637-1652.

Yang H, Wang H, Jaenisch R. 2014. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat Protoc.*

9(8):1956-1968.

Yoo CH, Na HJ, Lee DS, Heo SC, An Y, Cha J, Choi C, Kim JH, Park JC, Cho YS. 2013. Endothelial progenitor cells from human dental pulp-derived iPS cells as a therapeutic target for ischemic vascular diseases. *Biomaterials.* **34**(33):8149-8160.

Zhou T, Benda C, Dunzinger S, Huang Y, Ho JC, Yang J, Wang Y, Zhang Y, Zhuang Q, Li Y, Bao X, Tse HF, Grillari J, Grillari-Voglauer R, Pei D, Esteban MA. 2012. Generation of human induced pluripotent stem cells from urine samples. *Nat Protoc.* **7**(12):2080-2089.

Zhu H, Lau CH, Goh SL, Liang Q, Chen C, Du S, Phang RZ, Tay FC, Tan WK, Li Z, Tay JC, Fan W, Wang S. 2013. Baculoviral transduction facilitates TALEN-mediated targeted transgene integration and Cre/LoxP cassette exchange in human-induced pluripotent stem cells. *Nucleic Acids Res.* **41**(19):e180.

Zu Y, Tong X, Wang Z, Liu D, Pan R, Li Z, Hu Y, Luo Z, Huang P, Wu Q, Zhu Z, Zhang B, Lin S. 2013. TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat Methods.* **10**(4):329-331.

국문 초록

Transcription activator-like effector nucleases (TALEN) 과 RNA-guided engineered nuclease (RGEN) 은 최근에 도입된 유전체 공학의 도구로써 폭넓게 쓰이고 있다. TALEN은 DNA 염기 서열을 하나씩 특이적으로 인식하는 transcription activator-like effector protein 에 FokI cleavage domain 을 붙인 인공적인 제한효소이다. 본 연구에서는 *Pibf1* 과 *Sepw1* 유전자에 각각 특이적으로 작용할 수 있는 TALEN 을 제작하였고, 생쥐 세포주에 주입하였을 때 유전자에 변이가 성공적으로 도입되는 것을 확인할 수 있었다. 나아가 실제 생쥐의 수정란에 TALEN mRNA를 도입하였을 때 매우 효율적으로 녹아웃 된 생쥐를 얻을 수 있었다. RGEN은 세균이나 고세균에서 외부 도입 유전자를 절단하는 기능을 가진 CRISPR/Cas 시스템을 이용한 유전자 가위이다. 본 연구에서는 RGEN 을 사용하여 혈우병의 주요 원인으로 알려진 *F8* 유전자의 역위를 교정하는데 성공하였다. 먼저 Hela 세포주에서 *F8* 의 intron1 번과 22번에 작용하는 RGEN 을 각각 도입하였을 때 성공적으로 600kbp 에 이르는 염기 서열의 역위가 이루어진 것을 관찰하였다. 그리고 실제 혈우병을 가지고 있는 환자의 세포로부터 유도만능줄기세포(iPSC)를 얻은 후 해당 RGEN을 도입하여 *F8* 유전자의 mRNA 가 다시 정상적으로 발현되는 교정된 세포를 성공적으로 얻을 수 있었다. 이처럼 TALEN 과 RGEN 은 손쉽고 빠르게 제작할 수 있는 매우 효율적인 유전자 가위이므로 유전자의 특성 연구와 같은 기초 과학으로부터 유전자 교정이나 치료에 이르기까지 폭넓게 이용될 것으로 기대되고 있다.

학 번 : 2008-22716